Journal of Experimental Psychology

UNIVERSITY OF MICHIGAN

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

OCT 27 1955

SOCIAL SCIENCE LIBRARY

BF
1
J86

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Probability Discrimination in a Motor Task: L. B. WYCKOFF AND J. B. SIDOWSKI 225

Some Characteristics of Achievement Motivation: E. G. FRENCH 232

A Stochastic Model for Time-Ordered Dependencies in Continuous Scale Repetitive Judgments: B. WEISS, P. D. COLEMAN, AND R. F. GREEN 237

Self-Paced Rest with Variation in Work Loading and Duration of Practice: I. McD. BILODEAU 245

The Role of Extramaze Cues in Place and Response Learning: D. P. SCHARLOCK 249

The Facilitating Effect of Words on Discrimination and Generalization: D. DIETZE 255

Learning as a Function of Contexts Differentiated Through Antecedent Value Experience: E. GLOYE 261

The Effect of Change in Drive Level on Habit Reversal: A. M. BUCHWALD AND H. G. YAMAGUCHI 265

Wishing with Dice: R. A. MCCONNELL, R. J. SNOWDON, AND K. F. POWELL 269

Effect of Additional Spatial References on Display-Control Efficiency: W. D. GARVEY AND L. L. MITNICK 276

American Psychological Association

Vol. 50 No. 4 October 1955

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1955 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright © 1955 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 50, NO. 4 OCTOBER, 1955

PROBABILITY DISCRIMINATION IN A MOTOR TASK [1]

L. BENJAMIN WYCKOFF AND JOSEPH B. SIDOWSKI

University of Wisconsin

The study of the reactions of human *Ss*, when confronted with ambiguous situations, is of considerable theoretical and practical importance. One aspect of this problem has been investigated in a number of experiments referred to as verbal conditioning experiments, which have utilized a paradigm originated by Humphreys (**3**) and others (**2, 4**). In these experiments *S* was required to select one of two responses on each of a large number of trials. Each response was considered "correct" by *E* a certain percentage of the time in a random sequence, and *S* was given some signal to indicate correct responses. We will refer to the percentage of times that a given response was "correct" as the *yield* of that response.

In such situations *S* could maximize the number of correct responses by detecting which response had the higher yield and by making that response 100% of the time. One of the most striking results of these studies has been the finding that *Ss* do not adopt this optimal solution. Instead, they tend to make each response with a frequency approximately equal to its yield. We will refer to this mode of responding as a *matching solution*. This finding, if shown to be general, would have far-reaching implications both for learning theory and practical situations in which the consequences of a response are variable. The question may be raised whether the appearance of the matching solution is due to factors inherent in the operation of intermittent reinforcement or whether it might be due to certain other factors which these experiments have had in common.

All of these experiments have placed *S* in a situation where his assigned task has been to make a series of guesses or predictions. The responses to be made were essentially verbal in nature, requiring no skill. Recognition of this feature is illustrated by the use of the term "verbal conditioning" (**1, 2, 3**). Similarly the reinforcements have been simple cues which indicated that the response was correct or not correct. These characteristics appear to have set the stage for a kind of puzzle-solving

[1] This report represents part of the work of Contract No. AF 33(038)-23294 between the U. S. Air Force Personnel and Training Research Center and the University of Wisconsin.

attitude on the part of *S* that is illustrated by the typical attempts of *S* to find a pattern in the sequence.

The present study differed from the earlier studies in that both the guessing responses and the reinforcement were incorporated into a motor task. The situation was arranged so that guesses regarding an uncertain event would occur in the course of performing the task, even though *S*s were not instructed to guess and no direct reinforcement for guessing was used. It was hypothesized that under these conditions performance would tend to approach the optimal solution rather than the matching solution. Reasons for this hypothesis will be considered in the discussion.

The *S*'s task was that of tracking a target which swept across a screen in either of two directions in a random sequence. An attempt was made to arrange the task so that it would be very difficult unless *S* anticipated the target by moving the cursor to the correct side before the target actually appeared. Thus *S*s did engage in guessing behavior although this aspect of the task was not mentioned in the instructions.

The percentage of targets moving in each direction was varied systematically.

Method

Subjects.—The *S*s for this experiment were 60 male university students who were paid for serving as *S*s.

Apparatus.—The main components of the apparatus were (*a*) a display and hand-control unit, and (*b*) a programming and recording unit. A view of the display and hand-control unit is presented in Fig. 1. The target and cursor were presented as shadows on a milk-glass screen. The cursor consisted of a 1-in. open circle mounted on a 6-in. arm which could be rotated through a 90° arc by a 90° twisting motion of a wooden handle grip. The hand control had limit stops 45° on each side of vertical.

The target consisted of a ¼-in., solid circle mounted on a 6-in. arm which rotated through

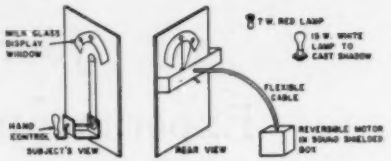

FIG. 1. The *S*s' view and rear view of display and control unit.

360° about the same axis as the cursor. The target arm was coupled to a reversible 15-rpm synchronous motor by means of a flexible cable. The motor was placed in a sound-shielded box to eliminate the possibility of differential sound cues which might otherwise indicate the direction of motion.

A 15-w. light 15 in. behind the screen cast a shadow of the target and cursor on the screen. A small red light, also behind the screen, could be lighted as a signal.

An electrical contact (not shown) mounted on the cursor hub and activated by a cam on the target hub provided an electrical "on target" signal whenever the target and cursor were within ½ in. of each other. Additional contacts indicated whether the cursor was in the right or left half of the field and provided other signals for the programming circuits.

The display and control unit was mounted in a window in the wall of a booth. The height of the hand control and display window were 30 in. and 50 in. so that *S* could work conveniently while seated in the office armchair which was provided. The booth was lighted with a 40-w. lamp placed so as not to shine directly on the display window. A speaker in the booth was used to present signal tones.

The programming and recording unit provided for completely automatic presentation and scoring of a sequence of trials, as indicated in the procedure below.

Procedure on individual trials.—On each trial the target started from the bottom position, moving clockwise or counterclockwise. The target appeared in the window 1.5 sec. later (having moved 135°), swept across the window in 1 sec., and continued to the bottom position. A buzzer which was activated at the same time as the motor served as a 1.5-sec. warning signal. At the end of one revolution the motor and buzzer stopped. The next trial started automatically 3 sec. later. The direction of motion was determined by an automatic device utilizing a punched tape. Targets appeared on the right or left in a random sequence which repeated after 200 trials. No other restrictions were imposed on the randomization, but the percentage of

targets moving from right to left varied as indicated in the experimental design. The same sequence was used for all *Ss* in a given group.

If the cumulative time-on-target during any one attack exceeded a preset interval, the red "hit" light was immediately lighted and a low-pitched tone was sounded in the booth, indicating a "hit." This circuit sometimes was disconnected under conditions described below.

If the cursor was within 1 in. of the center at the moment preceding the appearance of the target, the possibility of scoring a hit on that trial was canceled so that *Ss* would not adopt the habit of holding the cursor in the center position between trials. At the same time the "anticipation scoring circuit" was momentarily activated. If the cursor was on the right side of the center point, whether near the center or not, a right-side anticipation was recorded.

The number of hits and the number of right-side anticipations were recorded on a printing counter at the end of every 10 trials. Anticipations were recorded on only 9 out of each 10 trials beause of technical considerations.

Design.—The 60 *Ss* were randomly assigned to five groups of 12 *Ss* each who received 25, 40, 50, 60, and 75% targets from the right. Each of the five groups was divided into two subgroups which differed with respect to the amount of time-on-target required for a hit. The intervals used were .3 and .5 sec. A methodological problem was involved in picking an appropriate interval; thus, these two intervals were used.

Procedure.—The *Ss* were seated before the apparatus and given the following instructions:

"This is a tracking apparatus. This handle (demonstrating) controls the right and left movement of the tracking arm which you see through the glass. Note the large circle mounted atop the arm. When the experiment begins, a moving target will appear in the window. This target will consist of a rod with a dot upon it. Your job is to track the small dot, attempting to keep it within the larger circle of the tracking arm. If you can stay on target a certain length of time, a red light will go on, and a tone will be presented, and you will score one point. Try to score as many points as you can. Continue to track as well as you can until you are told to stop. Any questions?"

Each *S* was run for 320 trials, which took about 40 min., receiving the prescribed proportion of targets from the right. The "hit" timer was set to require .3 or .5 sec. on target for a "hit."

Results

Each trial was tabulated as either a right or left anticipation, depending on whether the cursor was on the right or left side of the center point. The percentage of correct anticipations in blocks of 20 trials was computed for each *S*. For this purpose we define a right-side anticipation as correct for the 60% and 75% groups, and left as correct for the 25% and

TABLE 1

Percentage of Correct Anticipations in Blocks of 20 Trials

Groups	Blocks															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
.3-Sec. Subgroups																
25%	63	65	72	84	66	88	89	93	93	87	88	83	79	86	84	88
40%	45	45	49	48	48	69	65	69	73	61	59	54	54	50	61	60
50%	40	31	30	31	28	22	26	22	27	28	19	26	19	15	24	26
60%	37	68	62	64	59	63	69	67	68	69	65	56	61	62	69	75
75%	54	66	64	79	71	67	65	71	71	75	67	64	69	74	73	74
.5-Sec. Subgroups																
25%	60	65	67	69	72	84	84	87	83	92	93	94	95	97	97	96
40%	43	45	39	41	45	62	62	61	68	77	72	69	77	86	71	78
50%	45	55	53	47	53	38	52	48	60	36	33	52	50	49	41	50
60%	46	52	61	71	80	73	85	66	62	81	81	75	78	87	83	77
75%	57	62	72	69	78	76	84	88	94	94	91	89	95	92	96	91

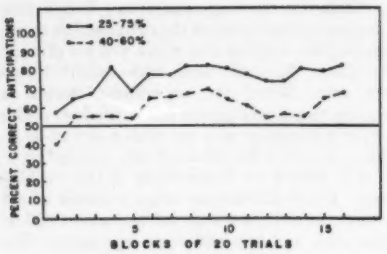

FIG. 2. Group mean learning curves for .3-sec. required-time-on-target subgroups

40% groups. Right-side anticipations were arbitrarily considered correct for the 50% groups. Group means are presented in Table 1. The 25% and 75% groups were combined in order to balance the effects of side preferences, as were the 40% and 60% groups. The combined group data for the .3-sec. subgroups are presented in Fig. 2. The corresponding curves for the .5-sec. subgroups are shown in Fig. 3.

The percentage of correct anticipations in the last 120 trials was taken as the terminal level for each *S*. These data are presented in Table 2. The differences between obtained terminal levels and the terminal level which would be expected on the hypothesis of a matching solution were tested by means of a *t* test (5) for an exact hypothesis. The results are presented in Table 3. For the

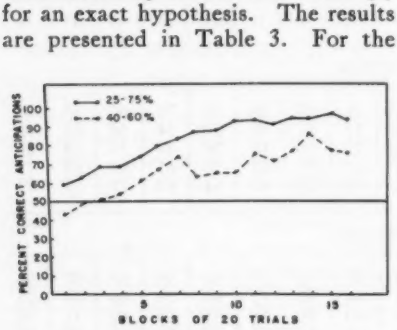

FIG. 3. Group learning curves for .5-sec. required-time-on-target subgroups.

TABLE 2

TERMINAL LEVEL IN PERCENTAGE CORRECT ANTICIPATIONS FOR LAST 120 TRIALS

Subgroup	Group				
	25%	75%	40%	60%	50%
.3 sec.	100	100	100	100	52
	100	100	100	99	30
	100	86	99	87	19
	93	77	19	52	15
	15	58	18	46	14
	40	0	0	5	0
.5 sec.	100	99	88	97	77
	97	98	86	93	61
	93	97	85	82	57
	87	97	81	82	46
	87	96	63	69	30
	80	86	60	56	4

TABLE 3

DIFFERENCES BETWEEN OBTAINED TERMINAL LEVELS AND TERMINAL LEVEL EXPECTED ON HYPOTHESIS OF A MATCHING SOLUTION

Sub-group	Comparison	Means	*t*	*P*
.5 Sec.	Obtained terminal level 25–75% group and expected 75%	93.08/75.00	9.72	.01
	Obtained terminal level 40–60% group and expected 60%	78.50/60.00	4.82	.01
.3 Sec.	Obtained terminal level 25–75% group and expected 75%	72.42/75.00	.25	
	Obtained terminal level 40–60% group and expected 60%	60.42/60.00	.04	

.5-sec. subgroups the terminal levels were significantly higher than the matching solution level beyond the .01 level of significance. No significant differences were obtained for the .3-sec. subgroups.

The number of "hits" were recorded primarily for the purpose of providing a check on the difficulty of the task for *S*s. These data will not be reported since they are of only secondary interest. We may, however, note the following: The task was neither unduly easy nor unduly dif-

ficult for the .5-sec. subgroups. All *Ss* succeeded in scoring frequently but none approached perfect performance. On the other hand, the task seems to have been somewhat too easy for the .3-sec. subgroups. Many *Ss* showed near-perfect scores toward the end of the session, and all of the .3-sec. subgroups showed mean frequencies of hits above the frequency of targets from the most frequent side, in all blocks from Block 6 on.

Some of the aspects of the results to be noted are as follows:

1. The group curves for the .5-sec. subgroups give no indication of a tendency toward a matching solution (a mode of behavior in which the performance percentages match the target percentages). On the contrary, the percentages of correct anticipations clearly continue to rise until they are well above the target percentages. Furthermore, these curves would not seem to indicate that asymptotic performance had been reached.

2. The group curves for the .3-sec. subgroups taken by themselves would seem to be quite consistent with the hypothesis of a matching solution. Group terminal performance, as measured by the last 120 trials, yields values very close to what would be expected on the basis of a matching solution. However, this observation must be tempered by consideration of the distributions of terminal levels. These distributions do *not* indicate that the group means accurately reflect individual performance. This point will be considered in more detail in the discussion.

Discussion

We will consider first some of the implications of the learning curves for the .5-sec. subgroups and then return to a discussion of the .3-sec. subgroups. Perhaps the most interesting finding in the present experiment is that, under certain conditions, performance curves continued to rise well above the level which would correspond to a matching solution, and moved in the direction of an optimal solution. This lends support to the notion that the appearance of a matching solution in previous studies (**2, 3**) was due, in part, to special features of the situations not inextricably associated with intermittent reinforcement.

One of the main differences between the present and previous studies (**1, 2, 3, 7**) lies in the fact that the present experimental situation placed the guessing aspect of the task in a secondary, inconspicuous position. The *Ss* were not told to guess nor were they ever given an indication that *E* was scoring them on their guesses (anticipatory responses). The reinforcement for anticipating the target lay in the direct advantage gained by *S* in his efforts to track the target.

In one sense, the reinforcement for anticipating was indirect in that points were not awarded for anticipating correctly but rather for tracking well. Thus, one might expect that the reinforcement would be less effective in the present case and that a high level of performance would be less likely to occur. However, it is not necessary to assume that the presentation of a "correct" signal has an equivalent effect in all experiments. We must not lose sight of the fact that the reinforcing effect of a "correct" signal is necessarily dependent on past experience in a rather complicated way. The effectiveness of such a signal in a particular situation probably depends on the significance which has been attached to such signals *in similar situations* in the past. Because of this, *Ss*' attitudes or reactions to the situation in general become very important, since they establish the generalizations which in turn control the effectiveness of the reinforcement.

More specifically, we may hypothesize that situations which place emphasis on guessing and prediction are closely related to past experience with certain kinds of games and puzzles. This re-

lationship carries two important implications: (*a*) When working a puzzle, intermittent "correct" and "incorrect" responses in an apparently meaningless pattern usually signify *failure*. Success is identified with *consistent* reinforcement. (*b*) In games where one is attempting to outwit or outguess an opponent, it is usually advantageous to change responses frequently so that your opponent can not predict *your* responses.

Another way in which past experience with games may control the reinforcing effect of the "correct" signal has been suggested by Grant (2), namely, many games place a premium on predicting improbable events or, in other words, "calling long shots." This factor would imply that the same signal might have a stronger reinforcing effect on responses with a low yield than on responses with a high yield. (Of course, the signal would not be obtained as often on low yield responses.)

All of the above factors would reduce the tendency to approach the optimal solution. This analysis does not offer any immediate explanation for the appearance of the matching solution, but it can account for the failure to obtain an optimal solution.

The present experiment, in contrast to the previous experiments, confronts *S* with a task involving motor skill and coordination. Neither the instructions nor the task itself suggest that there is a problem or riddle to solve. This situation is related to past experience with manual tasks and games involving skill. In such activities it is rarely required that a perfect response be made on every attempt. On the contrary, games of skill characteristically allow for indefinite improvement and no completely "correct" response is defined. Here failure to obtain consistent reinforcement is not taken as an indication of over-all failure.

Furthermore, motor-skill games generally do not emphasize the need to outwit an opponent. In throwing darts, for example, there is no need to be unpredictable. To be sure, it is advantageous to be unpredictable in some games such as tennis, but even here this aspect is probably less important than hitting the ball properly.

Thus we see that the factors which might tend to prevent the attainment of an optimal solution in studies emphasizing guessing, or suggesting a riddle to be solved, are absent or very much reduced if the guessing aspect is made subordinate to some other skill.

This analysis also suggests certain other conditions under which one might expect to obtain performance approaching an optimal solution. It might be possible to influence *S*'s attitude by means of instructions in such a way that generalization to previous experience with games would be reduced. In an experiment by Morin (6) some *S*s were told that one response was correct for each of several stimuli, but that the "correct" signal would not always be accurate. The *S*s were told, in effect, that if they failed to obtain consistent reinforcement, this might be the fault of the apparatus rather than theirs. These instructions would tend to reduce generalization to puzzle-solving situations in which lack of consistent reinforcement is an indication of failure. Morin's experiment was also complicated by the fact that eight different stimuli were used and the yield for the two responses depended on the stimulus presented. Results showed performance well above the matching solution level.

Findings for the .3-sec. subgroups.—It was observed, in Fig. 2, that group mean curves for the .3-sec. subgroups appeared to approach a matching solution. However, inspection of the terminal levels, in Table 2, for these *S*s indicates that the group mean curves do not reflect individual performance accurately. Virtually no individual *S*s reached a terminal level corresponding to the target frequencies. For example, in the combined 25–75% group, only 1 out of 12 *S*s reached a terminal level between 65% and 85%. This could hardly be taken as support for the prediction of a matching solution. One explanation for these findings is based on the fact that it was possible for *S*s to score hits even though they anticipated the target in-

correctly. The fact that this did occur was confirmed by *E*s' observations and also by the high frequency of "hits" recorded.

In view of this fact we conclude that the contingency between anticipation of the target and the opportunity to score was broken when *S*s attained a certain level of proficiency at the tracking task. Thus, after a certain point it made little difference to *S*s whether they anticipated the target correctly or not. Apparently at this point some *S*s had already acquired a tendency to anticipate on the most frequently correct side and simply continued with this pattern since there was no reason to change. Others had acquired no preference and continued to alternate from one side to the other. It is interesting to note that several *S*s acquired a tendency to anticipate on the least frequently correct side and continued with this pattern. We may conclude that the appearance of a group curve resembling a matching solution was a coincidence.

Summary

The present study tested the generality of certain findings of some verbal conditioning studies on the effects of intermittent reinforcement on human performance. The *S*s were forced to make a prediction regarding the path of a target if they were to succeed in the tracking task which was assigned them. In contrast to previous studies, *S*s were given no instructions regarding the problem of prediction nor was it indicated in any way that some predictions were considered correct and others incorrect by *E*. This arrangement was expected to reduce the tendency of *S*s to interpret the task as a guessing or problem-solving task.

The targets swept across a screen in an arc in either direction in a random sequence with a fixed percentage from each side. Percentages of 25, 40, 50, 60, and 75% were used for different groups. The *S*s were instructed to track the target with a cursor operated by a simple hand control. Each trial lasted 1 sec. and if *S* succeeded in staying on target for a cumulative total of .3 sec. (or .5 sec.), a signal was presented indicating a "hit." It was intended that the task would be very difficult unless *S* anticipated the target by slewing to the correct side before the target appeared. This intention was fulfilled for the .5-sec. groups but not for the .3-sec. groups.

The results for the .5-sec. groups show a gradual negatively accelerated increase in the percentage of anticipations toward the most frequently correct side. A level of well over 75% for the 25% and 75% groups and over 60% for the 40% and 60% groups was reached within the 320 trials of the experiment. This finding contrasts with the findings of several previous studies in which the response percentages approached the stimulus percentages. The .3-sec. groups showed considerably less improvement. This difference probably can be attributed to the fact that *S*s learned to score on the tracking task without anticipating the target correctly.

Reasons for the difference between the present results and previous results are considered in detail. Our interpretation includes the notion that the decreased tendency to interpret the task as a guessing or problem-solving task eliminates certain factors which otherwise interfere with the approach to an optimal solution. These factors are related to past experience with guessing and problem solving.

References

1. Estes, W. K., & Straughn, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
2. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
3. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
4. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
5. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
6. Morin, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, **49**, 243–251.

(Received November 22, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

SOME CHARACTERISTICS OF ACHIEVEMENT MOTIVATION [1]

ELIZABETH G. FRENCH

Personnel Research Laboratory, Air Force Personnel and Training Research Center

This study represents an attempt to investigate systematically: (*a*) the relationship between the administration of verbal instructions designed to vary achievement motivation and changes in scores on an independent measure of motivation, and (*b*) the relation of the first two variables, motivating conditions and measured motivation, to performance on a simple task. Since it is apparent that all individuals do not react in a uniform way to attempts to vary motivation by means of verbal instructions, a variable, or variables, in addition to the instructions must be operating. An hypothesis advanced here is that the individual's typical level of motivation is one of those variables. It is suggested that those *S*s who fail to respond to cues such as verbal instructions designed to raise achievement motivation are those with generally low need for achievement. Thus the typical level of achievement need, as well as the stimuli present designed to vary need level, must be taken into account in predicting motivation level in a given situation.

Two related hypotheses follow from this line of reasoning. One is that a greater amount of the variance in performance which is mainly a function of motivation rather than of skill will be due to this resultant total motivation than to environmental cues provided to raise the level of motivation. The second is that the degree of relationship between motivation level in one situation and both motivation level and performance level in another will depend in part on the degree of similarity between the situations with respect to motivational cues. The present experiment was designed to test these hypotheses.

METHOD

Instruments.—The motivation measure used was a device developed by the author for the measurement of complex motivations. It involves an adaptation of McClelland's (2) motivation imagery scoring system to responses to verbal items similar to those of Sherriffs' "Intuition Questionnaire" (3). The items are single-sentence descriptions of behavior which are presented as typical of hypothetical individuals with whom *S*s can identify (i.e., "Tom always lets the 'other fellow' win."). The *S*s are instructed to "explain" the behavior. The assumption is that an *S* with high achievement motivation, for example, will project that motivation into his interpretation of another's behavior when the cause of that behavior is unclear and when an interpretation in achievement terms is not unreasonable (or even, sometimes, when it is). The work of McClelland and of Sherriffs and preliminary work of the author indicate that this approach to the measurement of motivation has validity. The present instrument is disguised as a "Test of Insight." Two equivalent forms have been developed.[2]

[1] The research was carried out as part of the United States Air Force Personnel and Training Research and Development program. The opinions expressed, however, are those of the author and are not to be construed as reflecting the views or endorsement of the Department of the Air Force. Portions of this paper were read at the 1954 meeting of the American Psychological Association.

[2] An article discussing the rationale behind the test, describing the test and scoring system, and presenting preliminary validation data has been prepared for publication. The directions for the test, the test, and the scoring system have been deposited with the American Documentation Institute, and may be obtained by ordering Document No. 4607, remitting $1.25 for microfilm, or $1.25 for photocopies.

The performance test was a digit-letter substitution test which was very simple and primarily a speed test. The number of items completed could be assumed to be, at least in part, a function of the amount of effort expended. The test was of such a length that it could not be completed in the 5-min. time limit allotted. The *Ss* were instructed to work for both speed and accuracy.

Subjects.—The *Ss* were 90 male students in the Officer Candidate School at Lackland Air Force Base.[3]

Procedure.—Two experimental sessions were involved. In the first, regular Air Force test administrators gave all *Ss* one form of the Test of Insight during a regularly scheduled testing program which all candidates undergo at the start of the course. The scores ranged from 0 to 12 with a mean of 5.30 and *SD* of 2.02. The median was 4.5. The group was then dichotomized at the median of the achievement motivation scores and three experimental groups, each containing 15 high- and 15 low-scoring *Ss*, were formed. The value of *F* for the between-group variance in motivation scores was .13.

In the second session, which took place approximately five months after the first, *Ss* took one form of the digit-letter code test with no orientation other than instructions pertinent to the test itself, in order to establish a control performance level for each *S*. Then the verbal instructions designed to vary achievement motivation level were introduced.

The *Ss* in the first group, designated the relaxed group, were told by *E* (the author) in a casual and friendly way while she collected the code test papers:

"We are just experimenting today and we appreciate your cooperation very much. We want to find out what kinds of scores people make on these tests. The next test you are going to take you will recognize as being like one you took when you first came to OCS. We're asking you to take this second form today to try out the new items. After that you will take one more code test. If anyone has any questions, I'll be glad to answer them at the end of the period."

The second or task-motivated group was told the following, in a formal and serious manner, by a male airman after the code tests had been collected with obvious care:

"The test which you have just taken measures a critical ability—the ability to deal quickly and accurately with unfamiliar material. It is related to general intelligence, and will be related to your future career. Each man should try to perform as well as possible. In a little while you will be given a chance to better your score. But first, you will take a second form of an experimental test you took several months ago."

The third or extrinsically motivated group was given these instructions by a male airman after the code tests had been collected.

"We want to see how fast it is possible to work on a code test like the one you have just taken, without making errors. You are familiar with it now and know just what has to be done. Next you will spend 25 min. on a second form of a test you took several months ago. Then you will take a second 5-min. code test. The five men who make the best scores in that time will be allowed to leave right away—as soon as I can check the papers. The others will have more practice periods and more tests."

The total time scheduled for this group was so arranged that it would be immediately apparent to *Ss* that the five men who would be allowed to leave early would have an hour of free time, not an inconsiderable incentive for an OCS student.

The second form of the Test of Insight was given immediately following the instructions. When all *Ss* had finished (about 25 min.) the second code test was given. The *E*'s manner was in keeping with the particular instructions.

The Test of Insight papers were alphabetized without regard for experimental group and scored by the author. Unfortunately, a second scorer, who had assisted in the development of the test, was unavailable for this phase. In view of the high scorer agreement found (.88 and .91 for successive samples), however, it was felt that scoring could be done safely by one person. The scorer, of course, had no knowledge of the group from which any paper came.

Results

The main hypothesis stated that motivation under particular environmental conditions was a function of the typical motivation level of the individual as well as the cues in the environment introduced to vary motivation level. Tables 1 and 2 include the data relevant to this hypothesis. The mean achievement motivation scores for the six subgroups (two levels of initial motivation × three experimental conditions) appear in Table 1. Bartlett's test for homogeneity of

[3] The author acknowledges with sincere appreciation the assistance of Dr. Walter Borg of the USAF Officer's Military School in obtaining and scheduling *Ss*.

TABLE 1

MEAN ACHIEVEMENT MOTIVATION AND PERFORMANCE SCORES AS A FUNCTION OF INITIAL MOTIVATION LEVEL AND EXPERIMENTAL CONDITIONS

Initial Motivation	Experimental Conditions											
	Relaxed				Task = Ego				Extrinsic			
	Motivation		Performance		Motivation		Performance		Motivation		Performance	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
High	4.20	1.68	17.73	16.59	6.27	2.07	29.80	16.30	5.40	2.56	18.20	15.51
Low	2.80	1.34	15.40	17.70	4.40	2.18	16.60	14.20	5.67	1.39	22.47	14.33
r*	.17		.03		.45**		.48**		−.06		.02	

* Correlations with initial motivation.
** $P < .01$.

variance indicated that analysis of variance was not inappropriate for these data. The analysis (Table 2) shows initial achievement motivation level significant at the 5% level, as well as verbal instructions at the 1% level. While motivation level for the groups as a whole was affected by the experimental conditions, efforts to increase or decrease motivation did not override initial levels. Those who had high scores on the first test tended to have high scores on the second test and vice versa. Lack of significant interaction between the experimental conditions and initial motivation level suggests that the apparent reversal for the extrinsic group is a result of chance fluctuation.

The first related hypothesis stated that a greater amount of performance variance would be due to motivation as measured in the situation than to the environmental cues per se. Data in support of this hypothesis are presented in Tables 3 and 4. Perform-

TABLE 2

ANALYSIS OF VARIANCE OF SECOND ACHIEVEMENT MOTIVATION SCORES

Source of Variance	*df*	*MS*	*F*
Initial level of achievement motivation	1	22.51	5.25*
Experimental conditions	2	37.68	8.78**
Interaction	2	9.43	2.20
Within (error)	84	4.29	

* $P = .05$.
** $P = .01$.

TABLE 3

PERFORMANCE IMPROVEMENT MEANS AS A FUNCTION OF SECOND ACHIEVEMENT MOTIVATION SCORES

Motivation	Experimental Conditions					
	Relaxed		Task		Extrinsic	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
High	16.60	17.58	29.53	16.58	26.40	14.98
Low	16.60	16.75	16.93	13.51	14.27	12.61
r	.00		.48**		.51**	

* Correlation between motivation and performance.
** $P < .01$.

TABLE 4

ANALYSIS OF VARIANCE OF PERFORMANCE IMPROVEMENT SCORES WHEN SECOND ACHIEVEMENT MOTIVATION SCORE IS A VARIABLE

Source of Variance	*df*	*MS*	*F*
Achievement motivation scores	1	1529.34	5.92*
Experimental conditions	2	331.74	1.28
Interaction	2	382.75	1.48
Within groups (error)	84	258.36	

* $P < .05$.

ance improvement scores were obtained by subtracting each *S*'s first code test score from his second. New subgroups of 15 *S*s each were formed by dichotomizing the experimental groups at the median of the Test of Insight scores obtained under the experimental conditions. The means for performance improvement for these groups appear in Table 3. The analysis of variance (Table 4) shows that the *F* for achievement motivation scores is significant at the 5% level. That for experimental conditions does not reach conventional standards of significance. Motivation, as measured in the situation, did prove to be more closely related to task performance than did the motivational set created by instructions alone. In fact, had the experimental conditions been the only measure of motivation used, it would have been necessary to conclude that achievement motivation did not affect the task performance even though we know that, due to the nature of the task, this is highly unlikely.

The last hypothesis was concerned with the degree of relation between motivation level in one situation and motivation level and performance in another as a function of the similarity of the motivation cues in the two situations.

It should be pointed out here that the original design of the experiment called for the administration of the first motivation test as an experimental test on a day when other experimental tests were being given, in order to minimize achievement cues. However, it actually was given along with the Career Counselling Battery which the men knew was important for their future assignments. This was a real situation similar to the one the task-motivated directions tried to create. Given these conditions, the expectation was that the relation between original motivation level and both final motivation level and performance scores would be greatest in the task-motivated group, somewhat less in the relaxed group where no achievement cues were provided by the environment, and least in the extrinsically motivated where an entirely different set of motivational cues was introduced. The interaction value in Table 2 and a corresponding one computed for the performance data in Table 1 provide the over-all test of this hypothesis. The *F* ratio of 2.20 found in Table 2 and the one of 2.16 obtained for the performance data do not reach the .05 level of significance. They are large enough to be suggestive, however, and the differences in the table are in the predicted direction. Therefore, correlations between original motivation scores and the other sets of scores were obtained to provide more information. These are presented in Table 1. Those for the task-motivated group are both significant beyond the .01 level; those for the other groups do not reach the .05 level. In addition, according to expectation, both correlations for the task-motivated group are significantly greater than the corresponding ones for the extrinsically motivated group. The difference between the correlations for the task-motivated and relaxed groups is significant for the performance scores but not for the motivation test scores. These results give support to the hypothesis that similarity of motivation level from situation to situation is most likely to be found when environmental cues relevant to motivation are similar, and that performance can best be predicted under those circumstances as well.

Discussion

Only one point needs further discussion. Although the interaction given in Table 4 also proved to be nonsignificant, examination of the means in Table 3 raised a question concerning the relaxed group. Correlations computed between second achievement scores and performance scores separately for the three conditions yielded values for the two motivated groups that were significantly higher than the zero value obtained for the relaxed group. This was contrary to prediction. An explanation advanced here is that this finding could be due to the nature of the instructions to this group. The task was presented to *S*s in such a way that it might have appeared uninteresting to those with high achievement motivation but stimulating to those with high affiliative motivation. The achievement or goal-attainment aspect of the situation was de-emphasized by the instructions. But the female *E* was friendly in dealing with the group and asked for cooperation. Achieving high scores could be perceived as a way of pleasing *E*. Under these circumstances *S*s with high affiliative motivation would be expected to work hard on the task. Atkinson (1) has suggested such a possibility in the interpretation of his experimental results. This hypothesis was tested by rescoring the Test of Insight papers for affiliation imagery and correlating these scores with performance. These correlations were .48, .02, and − .13 for the relaxed, task, and extrinsic groups, respectively. The relationships expressed are the exact reverse of those for achievement motivation shown in Table 3. Apparently the relaxed situation was not only lacking in achievement motivation cues but contained cues relevant to affiliation as a motive.

In general our findings were in support of the hypotheses proposed: that what has been described as typical level of achievement motivation is a significant variable affecting the degree to which a desired level of motivation can be aroused by introducing appropriate cues into an experimental situation; and that this resultant motivation, as it was measured, was related to performance. The findings suggest that an independent measure of motivation and a knowledge of the characteristics of the stimulus situation are both essential for predicting performance. The incidental finding, not specifically predicted, that it is possible to some extent to identify in advance which *S*s are more likely to increase output under a given set of stimulus conditions, is of considerable interest.

The final point to be made is that the results, which are in accordance with the hypotheses proposed, provide additional validation for the Test of Insight as a method of measuring achievement motivation.

Summary

An independent measure of motivation and a performance test were given under three different verbally created conditions of achievement motivation: relaxed, task motivated, and extrinsically motivated.

The results, consistent with hypotheses proposed, may be summarized as follows:

1. Increase in achievement motivation score was a function of both previous motivational level and the experimental conditions.
2. Performance scores were more closely related to motivation scores than to the experimental conditions.
3. Performance scores in one situation tended to be most closely related to motivation scores in another when the situations presented similar motivational cues.
4. In addition, when affiliation cues were more prominent in the situation than achievement cues, performance was related to affiliation motivation scores rather than achievement motivation scores.

References

1. Atkinson, J. W. The achievement motive and the recall of interrupted tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.
2. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
3. Sherriffs, A. C. "The Intuition Questionnaire": a new projective test. *J. abnorm. soc. Psychol.*, 1948, **43**, 326–337.

(Received October 22, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

A STOCHASTIC MODEL FOR TIME-ORDERED DEPENDENCIES IN CONTINUOUS SCALE REPETITIVE JUDGMENTS [1, 2]

BERNARD WEISS,[3] PAUL D. COLEMAN,[4] AND RUSSEL F. GREEN

University of Rochester

The impetus for the study reported here arose from a phenomenon observed in our laboratory in the performance of a simple psychophysical task: bisecting a perceived angular distance by setting a knob on the basis of kinesthetic cues. This phenomenon takes the form of an initial drift in constant error. Typically, *Ss* tend to make their first setting close to the required value; on the settings that follow, they tend to overshoot more and more until they reach a range of values at which they begin to level off. When the data for all *Ss* are pooled, the initial drift seems to persist for about 3 to 10 trials. This initial drift stimulated the suspicion that a long time series of adjustments might yield an oscillatory function dominated by periods the length of the original drift; and so an investigation was planned to study the pattern of serial dependencies in such a time series.

Psychologists have been aware for some time that the successive responses of an *S* in many types of psychophysical situations are not independent. But, until lately, with the application of more sensitive and efficient techniques, the nature of the dependency was not clear. Some recent studies using these techniques have more successfully probed the problem. Autocorrelation analysis, for example, has been used to uncover any serial dependencies present in sequences of yes-no responses, such as those made in threshold or DL studies (**3, 7, 11, 12**). The general conclusion imposed by these and similar studies is that *S* tends to give a "yes" response if the previous response was "yes," and vice versa. Another technique suitable for such applications is spectral analysis. Abelson (**1**) tried to use this technique to describe individual differences in the performance of a simple repetitive task. Reliable individual differences appeared in the production of low-frequency components, indicating that in the output time series there were differences in the degree to which neighboring observations tended to be alike. Wertheimer (**12**) used spectral analysis in addition to autocorrelations to measure fluctuations in threshold measurements and found significant serial dependencies.

These techniques may also be used to help choose among alternative models to represent the underlying stochastic process in the time series. A stochastic process may be defined as a process in which one or more stochastic (random) variables enter

[1] This study was conducted as part of a research program on human motor performance supported by a contract between the Special Devices Center, Office of Naval Research, and the University of Rochester with Dr. S. D. S. Spragg as the responsible investigator. The authors would like to thank Dr. Spragg for his support and encouragement.

[2] The authors are indebted to Drs. John Tukey, Karl Lowy, Arthur Dutton, and William F. Taylor for suggestions and comments on the manuscript.

[3] Now at the USAF School of Aviation Medicine, Randolph Field, Texas.

[4] Now at the Army Medical Research Laboratories, Fort Knox, Kentucky.

whose probability distribution depends on a continuously varying parameter, usually time. The object of the present study was to obtain a stochastic model for the time-ordered dependencies in the bisection task.

Procedure

To secure time series of sufficient length on this task—none of the earlier *S*s had been run for more than 30 trials—20 *S*s were required to make 120 bisections of an angle on each of two consecutive days. The angle measured 40°. The *S*, who was always blindfolded and never received knowledge of results, sampled the angle (whose limits were set mechanically) by rotating the knob twice through it; he then tried to bisect the angle. To insure against a loss in motivation, *S*s were promised bonuses for accuracy: 5 cents for each setting within 1° of center, 1 cent for each setting within 2° of center. The *S*s were paid at the conclusion of the experiment. Apparently this procedure succeeded in maintaining the original level of motivation since variability did not show a significant rise.

While frequency of adjustment was not uniform, the time between the termination of one trial and the onset of the next was almost constant (about 5 sec.); *S* was self-paced during the actual setting. The *S*, it should be noted, since he was judging on a continuous scale, could choose from a theoretically infinite set of possibilities.

Results

Autocorrelation and spectral analysis.—The two chief methods of analysis used were autocorrelation and spectral analysis. These are believed to be the most powerful methods available for revealing serial dependencies and are capable of detecting significant periodicities in the time series if they are present.

These methods of analysis assume stationary time series. A stationary time series is one whose basic characteristics (probabilities) do not vary when it is shifted along the time axis. That is, the distributions of the random variables generated by the stochastic process are all the same and do not depend on time. Thus, if the original series is displaced by q observations and the properties of the series do not change, it is said to be stationary. Because our data revealed no obvious trends after the initial drift, we considered it legitimate to assume that the series obtained were stationary.

Since autocorrelation analysis involves laborious computations, we first tried to discover whether the time series generated by the *S*s were grossly random. For this purpose, the modification of the Wald-Wolfowitz test given by Moses (5) was used, and the results are shown in Table 1. A run is defined as a series of consecutive settings above or below the median. From the number of significant runs obtained it was evident that the *S*s did not generate random series (6) and so it was decided to employ the greater descriptive sensitivity of autocorrelation analysis.

An autocorrelation is computed in the same way as an ordinary product-moment correlation coefficient, except that it is used to reveal the relationship among observations separated by a certain number of intervening observations. Thus, the autocorrelation of Lag 1, R_1, is the correlation co-

TABLE 1

Number of Runs Made by Each *S* for First and Second Session

(T = Number of consecutive settings above or below the median)

S	T_1	T_2	*S*	T_1	T_2
1	30*	41*	11	49*	40*
2	54	54	12	51	28*
3	42*	54	13	42*	49*
4	57	46*	14	44*	48*
5	35*	43*	15	45*	56
6	51	50*	16	47*	48*
7	37*	48*	17	42*	54
8	45*	52	18	49*	45*
9	50*	46*	19	42*	58
10	53	61	20	59	52

* Significant at 5% level or less (Wald-Wolfowitz test).

TABLE 2

AUTOCORRELATIONS OF LAGS 1–5, 10, AND 20, FOR 10 *Ss* OVER THE FIRST 120 TRIALS

Ss	Lag						
	1	2	3	4	5	10	20
2	.37	.27	.24	.13	.18	.01	−.04
5	.43	.21	.15	.05	−.04	.31	.12
7	.27	.36	.24	.09	.24	−.15	.01
9	.25	.13	.05	.11	.01	−.06	−.22
13	.25	.28	.32	.23	.11	−.07	−.16
15	.43	.32	.18	.17	.19	−.07	.03
16	.34	.17	.06	.02	.13	.13	.08
18	.46	.43	.32	.38	.32	.29	.01
19	.57	.45	.32	.12	.08	.08	.04
20	.44	.31	.29	.25	.43	.21	.16
Mean*	.39	.30	.22	.16	.15	.07	.001

* Mean based on *z* conversion.

efficient between the series $X_1, X_2, \ldots, X_{n-1}$, and the series $X_2, X_3, \ldots, X_n$, where each pair of values consists of observations X_t and X_{t+1}. Autocorrelations of greater lags are obtained in a similar fashion. In the present study, autocorrelations of Lag 1, 2, 3, 4, 5, 10, and 20 were computed for 10 of the 20 *Ss* for the first 120 trials. These 10 were selected at random. The data for the other 10 *Ss* were not used in the computations because of considerations of time and expense. The picture that emerged, however, was quite clear, as can be seen in Table 2. The coefficients for the group as a whole, obtained via the *z* transformation, fall steadily in size as lag increases. A specific setting, therefore, tends to be most related to the preceding one, with influence declining as the temporal relationship becomes more remote. This pattern of dependency also indicates that the series are subject to long-period drifts and fluctuations. Such a pattern, of course, may or may not wholly fit the series generated by the individual *Ss*. The correlogram, or plot of autocorrelation coefficients against lag, is shown in Fig. 1.

The relationships suggested by the correlogram, however, are obscured in many cases by the sampling deficiencies of autocorrelation analysis. These arise from the lack of independence among the coefficients. If, for example, the autocorrelation of Lag 1 is close to unity, then the autocorrelation of Lag 2 must also be high.

The technique known as spectral analysis avoids these sampling deficiencies by postulating a different and more complete model. Its fundamental idea is that any stationary time series can be described as a sum of an infinite number of cosine-wave components of particular frequencies and amplitudes. This will be recognized as a statement based on the class of theorems originated by Fourier. Tukey (**9, 10**) has provided this model with certain useful statistical properties by assuming that phase variations occur randomly. This proposition can thus be expressed[5] as

$$X_t = \sum_{i=1}^{\infty} a_i \cos(\omega_i t + \Theta_i)$$

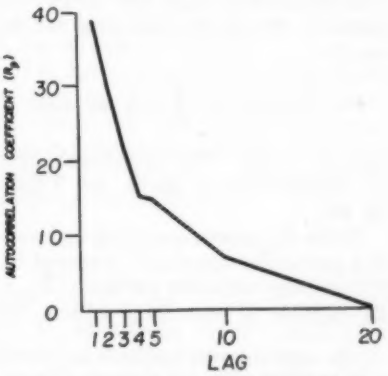

FIG. 1. Correlogram based on combined autocorrelation coefficients for 10 *Ss*.

[5] The derivation is given by Abelson (**1**), who also provides a clear and detailed exposition of this method.

where X_t = random variable observed at time t, ω_i = angular frequency (fixed), a_i = amplitude (fixed), and Θ_i = phase. Θ_i varies randomly and contributes the variability among time-series samples.

Such a time series has variance at angular frequency ω_i equal to the value of a_i^2. The relationship between these amplitude values and the total variance of the time series is given by

$$\sigma_x^2 = \lim_{n\to\infty} \sum_{i=1}^{n} (1/2)\, a_i^2$$

where σ_x^2 = total variance.

If we plot the contribution to the total variance of each component frequency, we obtain a distribution of variance over frequency. Tukey says (**10**) that it is as natural, therefore, to speak of a *variance spectrum* as of a *power spectrum*. He thus conceives the purpose of spectral analysis to be to estimate how energy (power, variance) is distributed over frequency for a given time series. These estimates, known as the *spectral densities*, $s(\omega)$, are obtained from the lagged cross-products,[6] since the latter are linked to the spectral densities by the relation

$$\text{ave}\,(X_t X_{t+p}) = \int_0^{\pi} \cos p\omega s(\omega) d\omega$$

where p = lag. The actual technique of computation is given by Tukey (**9, 10**).

While the power spectrum estimate of a particular frequency, denoted by U_p, and the adjacent estimate, U_{p+1}, do, because of smoothing effects, influence one another, any U_p and U_{p+2} are wholly independent. This allows confidence limits for the estimated values to be computed without distortion by sampling errors in remote estimates, and thus permits a test of significance [7] of the spectrum, a procedure impossible to perform with the correlogram. Since the aim of the present study was to devise a stochastic model to represent the obtained data, power spectrum estimates were computed to test the fits of the models proposed; they were not computed to provide a spectral model.

The power spectrum is interpreted in the following way. If the spectral estimates increase in value as ω goes from its limits of 0 to π (provided the t's are unit increments), the time series tends to be dominated by high-frequency components and will reveal

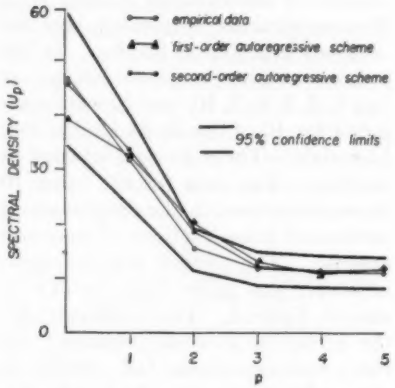

FIG. 2. Power spectrum estimates for the observed data and the first-order and second-order autoregressive schemes, with 95% confidence limits for the observed data.

[6] The spectral density function is the Fourier cosine transform of the autocorrelation function. Indeed, according to a theorem by Wald (see **2**), the necessary and sufficient condition that the autocorrelation exist is that $S(\omega)$, where $dS(\omega) = s(\omega)d\omega$, be a distribution function defined between 0 and π.

[7] The test of significance is based on the property that $(h\, U_p/\hat{U}_p)$ is approximately distributed as χ^2 with h degrees of freedom, where $h = 2N/m$ and $\hat{U}_p$ = population value of U_p, the power spectrum estimate. (N = number of observations, m = number of lag covariances used in the computations.)

TABLE 3

AUTOCORRELATIONS AND POWER SPECTRUM ESTIMATES YIELDED BY THE EMPIRICAL DATA AND BY THE FIRST- AND SECOND-ORDER AUTOREGRESSIVE SCHEMES

(95% confidence limits for the empirical estimates)

Lag	Empirical Data		First-Order Autoregression Predictions		Second-Order Autoregression Predictions	
	Auto-correlation	Power Spectrum Estimate	Auto-correlation	Power Spectrum Estimate	Auto-correlation	Power Spectrum Estimate
0	1.00	47.23±12.43	1.00	39.45	1.00	45.73
1	.39	31.85± 8.32	.39	32.34	.39	33.74
2	.30	15.51± 4.05	.15	20.38	.29	18.66
3	.22	11.91± 3.11	.06	13.10	.16	12.27
4	.16	11.47± 2.99	.02	11.21	.10	11.36
5	.15	11.29± 2.95	.01	12.04	.06	11.99

quick ups and downs; if they decrease in the same direction, the series tends to be dominated by low-frequency components and to produce long wave-length cycles and drift.

Power-spectrum estimates for the present data were computed from the mean autocorrelations of Lags 1 through 5. These are plotted in Fig. 2 (empirical series) and given in Table 3 under the heading "empirical series." Confidence limits (95%) are also given. These estimates show the empirical series to be dominated by low-frequency components, which indicates, like the correlogram, that adjacent settings tend to be more related than remoter ones, and that the series are subject to long wave-length drift and fluctuation. The high-frequency components, which tend to produce rapid alternations, contribute only a small portion of the total variance.

The autoregressive scheme.—Although stationary time series may be generated in several ways, the one with which psychologists would typically be concerned is the autoregressive process. This is a series generated by the recurrence formula

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-k}) + \epsilon_t$$

where ϵ is an error term taken to vary randomly. Such a formulation seems to be the most useful to the psychologist because, customarily, it is concerned with the linear determination of any observation from the previous ones. Such an equation represents an oscillating system subject to random shocks, for example, an oscillating pendulum bombarded irregularly by small boys equipped with peashooters (**2**). It should also be observed that the autoregressive model contains a random *additive* error. (See Kendall, **4**, and Tintner, **8**, for further details.)

The first-order linear autoregressive scheme, which represents a time series in which only the previous observation is needed to make the best prediction of the next observation, is given by the expression

$$X_{t+1} = aX_t + \epsilon_{t+1}.$$

In the present case, the estimated value of $a = .39$. (See Tintner, **8**, for an exposition of the method of obtaining coefficients for the autoregressive model.) The theoretical autocorrelations of the model can be computed from the autoregression coefficients. For the first-order model, the pth lag autocorrelation, R_p, is

equal to a^p for long time series. This yields the values for Lags 1, 2, 3, 4, and 5 given in Table 3.

The second-order linear autoregression equation, which may be regarded as expressing the regression of X_{t+2} on X_{t+1} and X_t, is of the form

$$X_{t+2} + aX_{t+1} + bX_t = \epsilon_{t+2}.$$

The present data yield the estimated values $a = -.326$, $b = -.165$. The theoretical autocorrelations are given in Table 3.

The accuracy with which these models reflected the empirical series was estimated by computing power-spectrum estimates from the theoretical autocorrelations yielded by the first- and second-order autoregression equations and plotting them on the graph of the power spectrum of the empirical series. The power-spectrum estimates for all three series appear in Table 3 and are plotted in Fig. 2.

While the second-order model seems to yield a slightly better fit than the first-order model, the difference does not seem to us to be significant. On the principle of parsimony, therefore, the first-order model can be accepted as the one which describes the stochastic structure of the present series.[8] Furthermore, Dr. Tukey (personal communication) has proposed another first-order model, which, provided certain assumptions are made, fits the observed data even more closely. Consider the two series

$$Y_{t+1} = aY_t + \epsilon_t$$
$$X_{t+1} = Y_{t+1} + \eta_t$$

where Y_t is what S is attempting, and what he actually does, X_t, is compounded from Y_t and η_t. The ϵ's and η's are independent random terms. Then, if

$$\text{cov } (X_{t+p}, X_t) = \text{cov } (Y_{t+p}, Y_p),$$
$$\text{cov } (X_{t+p}, X_t) = a^p \sigma y^2$$
$$\text{var } (X_t) = \sigma y^2 + \sigma \eta^2$$
$$R_p = \left(\frac{\sigma y^2}{\sigma y^2 + \sigma \eta^2}\right) a^p.$$

With $a = .78$ and $\sigma y^2/\sigma \eta^2 = 1.00$, the theoretical autocorrelations yielded by this model differ only slightly from those yielded by the empirical series.

The first-order autoregressive scheme is also called a Markoff process. This is formally defined as a stochastic process in which the cumulative conditional probability distribution of the stochastic variable X_t, given the variables X_{t-1}, X_{t-2}, . . . , is equal to the cumulative conditional probability distribution, given only X_{t-1}. Thus

$$E\,(X_t | X_{t-1}, X_{t-2}, \ldots) = E\,(X_t | X_{t-1})$$

where E is the mathematical expectation (probability).

Thus, once the value of X at the last available value of t is known, no previous history of the series is needed to make the best prediction about the future development of the system. It follows that the state of the system and its development are independent of the way the system was started. In terms of the present experiment, this means that, knowing only the last setting he has made, we can predict S's future settings as precisely as the random variation will permit.

Discussion

The model yielded by these analyses has not provided a solution to the drift problem, although it disposes of the hypothesis that the series follow periodicities of the length of the original drift. There are more important implications, however, arising from the excel-

[8] The reader should again note that the power spectrum is not itself being used as a model, but, rather, as a way of testing different autoregressive models.

lent fit of the Markoff process model to our data. For we believe that most psychologists would have predicted a significant relationship of higher order than the first, i.e., that they would have presumed that knowledge of the past history of the organism plays a larger part in the prediction of performance than our results would indicate.

Thus, while the particular task we chose is not wholly comparable to many others which are used in psychophysics, we deem it legitimate to suggest that some alteration is required in the view that an individual's judgment on a particular trial is significantly influenced by a large portion, or even the entire history, of that individual's experiences on the task. The present data, to repeat, are fitted by a model in which the development of a system is governed by only its present state and in which the history of development is accorded no significant role.

For binary "yes-no" judgments, evidence is available which indicates that the serial dependencies in such tasks can also be described as Markoff processes. This is shown explicitly in Abelson's analysis of an unpublished experiment by Day (1) and by Flynn, Jerome, and Moody (3). The results adduced by Senders (7) and by Verplanck, Collier, and Cotton (12) also indicate that they might be described by the Markoff model. The significant autocorrelations obtained by Senders are heavily concentrated at Lag 1. The critical ratio plot given by Verplanck, Collier, and Cotton, which presumably reflects the size of the serial relationship, shows an exponential decline with increasing lag. Although these authors claim significant dependencies as far out as Lag 11, their test of significance was not entirely legitimate since it does not take full account of the influence of less remote relationships. Using the critical ratio values as a rough index of the autocorrelation coefficients, one may see that their data might be fitted by a first- or second-order autoregressive model.

Although the correspondence between these results and ours is gratifying, we must assert that they are not directly comparable. Since they were obtained at threshold, or when the stimuli were equated, the yes-no judgment dependencies probably represent guessing habits. Our results, we submit, reflect dependencies which can be attributed to what might be called a kinesthetic trace, and therefore possess implications for judgment models like adaptation-level theory.

Summary

Previous data showed that *Ss* performing a task in which they were required to bisect, on the basis of kinesthetic cues, a perceived angular extent (by turning a knob) tended to produce an initial drift in constant error. To study more clearly the serial dependencies in this task, 20 *Ss* were required to perform long series of these adjustments.

Preliminary analysis showed the time series generated in this manner to be nonrandom. To procure a more complete picture of the time-ordered dependencies in the data, the time series generated by 10 of the 20 *Ss* were subjected to autocorrelation and power spectrum analysis. These techniques revealed that neighboring observations tended to be most alike, with influence declining as the temporal separation became greater.

A stochastic model was found for this pattern of serial dependency which could be described as a Markoff (first-order autoregressive) process. This means that for the best prediction of a setting only the previous setting need be known, the history of earlier performance making no direct contribution to the prediction. On this basis, it is suggested that remote past experience in kinesthetic judgments is of less consequence than commonly assumed.

References

1. Abelson, R. P. Spectral analysis and the study of individual differences in the performance of routine, repetitive tasks. Princeton, N. J.: Educ. Test. Serv., 1953.
2. Bartlett, M. S. *Stochastic processes.* Chapel Hill, N. C.: Institute of Statistics, Univ. of North Carolina, 1947.
3. Flynn, J. P., Jerome, E. A., & Moody, J. A. The intraserial dependence of psychophysical responses. *Amer. Psychologist,* 1952, **7,** 281–282. (Abstract)
4. Kendall, M. G. *Contributions to the study of oscillatory time-series.* Cambridge: Cambridge Univ. Press, 1946.

5. Moses, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.
6. Sakoda, J. M., Cohen, B. H., & Beall, G. Tests of significance for a series of statistical tests. *Psychol. Bull.*, 1954, **51**, 172–175.
7. Senders, V. L. Further analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1953, **66**, 215–228.
8. Tintner, G. *Econometrics*. New York: Wiley, 1952.
9. Tukey, J. W., & Hamming, R. W. Measuring noise color. Bell Tel. Lab., Dec., 1949.
10. Tukey, J. W. Measuring noise color. Memorandum prepared for distribution at meetings of Metropolitan Section, I.R.E., Nov., 1951.
11. Wertheimer, M. An investigation of the "randomness" of threshold measurements. *J. exp. Psychol.*, 1953, **45**, 294–303.
12. Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, **42**, 273–282.

(Received October 30, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

SELF-PACED REST WITH VARIATION IN WORK LOADING AND DURATION OF PRACTICE[1]

INA McD. BILODEAU

Skill Components Research Laboratory, AF Personnel and Training Research Center

The effect of work loading on the performance of a simple cranking response has been investigated in a number of studies (**1, 2, 3**). The studies showed marked differences in performance *while* differences in loading prevailed, but when loading differences no longer obtained, clear-cut effects attributable to prior loading were difficult to demonstrate. The performance differences between groups shifted from different work loadings to a common loading were slight. Any one study offered very little ground for rejecting the hypothesis that inhibition was independent of work loading, although the data did indicate, fairly consistently, a slightly more negative effect from prior practice against the heavier of two loadings.

In the above studies, the postshift test was given under the loading used by one of the loading groups in preshift cranking. A habit alternative can thus be offered to account for the postshift differences obtained between groups working against the lighter of the two loadings employed: in effect, that *Ss* shifted from the heavier to the lighter loading suffer a habit disadvantage relative to *Ss* working against a loading with which they are familiar. One study in which bidirectional shifts in loading were made affords some support to the habit alternative, with slight differences favoring *Ss* who worked against the heavier loading in pre- and postshift practice over *Ss* who worked against the lighter loading before the shift (**1**). A second study revealed little difference in rate of responding against the heavier loading after either of two rests interpolated between pre- and postshift practice, though the longer rest yielded sizable differences between rates of responding against the lighter loading (**2**). The latter experiment suggests that both habit and inhibition may determine the postshift differences, acting together when the test is against the lighter loading and in opposite directions in tests against the heavier loading.

This paper presents a study in which the index of inhibition, self-paced rest, is without apparent habit bias. If there is a difference in the amount of inhibition accrued in work against different loadings, and if *Ss* allow sufficient rest for complete recovery, the difference should be reflected in the times required to dissipate the different amounts of inhibition. Unpublished data, collected by the writer, R. A. Goldbeck, and B. Reynolds on a self-paced learning task, indicated that self-paced resting data are correlated with the usual indices of differential inhibition, mean rest increasing as the duration of prerest practice increased. Different durations of prerest practice were used in the present study, also, in order to explore the interaction of work loading with trial duration and to provide a check, with loading con-

[1] This research was carried out under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, in support of Project 7707. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

stant, of the adequacy of self-paced rest as an index of inhibition.

Method

Subjects.—The *S*s were 400 basic trainee airmen supplied by Lackland Air Force Base and assigned to treatments according to successive, unsystematic testing sequences employing each treatment once.

Apparatus.—The apparatus was a manual crank (**1, 2, 3**) requiring *S* to rotate a handle against one of two work loadings. The loadings are called Loads 0 and 4, with Load 4 requiring about three times the horsepower required by Load 0 at any given rate of turning.

A switch connected to a clock on *E*'s panel was placed to the left of the apparatus on the table to which the crank was attached. Pressing the switch sounded the buzzer and stopped the rest clock. The rest clock was started by *E*.

Procedure.—There were 10 groups of 40 *S*s in a factorial design with five durations of prerest work and two work loadings. The *S*s cranked against Load 0 or Load 4 for 10, 30, 60, 120, or 240 sec., rested, and then cranked against Load 0 for 120 sec.

The interpolated rest was self paced, with *S* instructed to rest until he felt ready to do his best. The *S* started each of his two work periods by pressing the switch with his left hand and simultaneously starting to crank with his right hand. Within periods, *S* was told to turn the crank as fast as possible until *E* said "Stop." The *S* was told neither the duration nor number of his work periods, nor was he given any information about work loading. The instructions indicated that he was to have several trials, that in each trial he was to do his best, and that between trials he was to rest until he could do his best. Between work periods, *S* was allowed to remain standing or to take a seat behind a screen shielding him from the apparatus.

The measure of the self-paced rest was taken from the rest clock that started when *E* ended the prerest period and stopped when *S* pressed the starting switch. The performance measure recorded was rev./10 sec., taken from counters on *E*'s panel.

Results [2]

As in previous studies, rev./10 sec. in prerest practice distinguished clearly between groups working against different loadings, while within loadings, performance fell off regularly over successive scoring periods. In terms of rate of responding in the prerest cranking period, thus, the groups were differentiated according to both experimental variables.

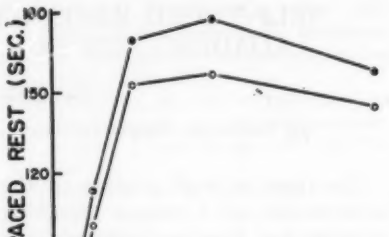

Fig. 1. Mean self-paced rest as a function of duration of prerest practice. Prerest work loading is the parameter.

In Fig. 1, mean self-paced rest in seconds is plotted against mean duration of prerest practice in seconds; work loading is the parameter. For either loading, the mean duration of self-paced rest increases with negative acceleration as prerest cranking increases up to 120 sec., with a decline in self-paced rest between the 120- and 240-sec. values of prerest work. At least for the first four durations, then, self-paced rest seems to agree in direction with other measures used as indices of the relative amounts of inhibition accrued with increasing durations of prerest practice.

It is also apparent that, for each

[2] Tables of data presenting for each group the means and variances of (*a*) the no. rev./10 sec. in the first and last trials of pre-and postrest cranking, and (*b*) the self-paced rest have been deposited with the American Documentation Institute. Order Document No. 4609, remitting $1.25 for microfilm or $1.25 for photocopies.

duration of prerest cranking, the mean self-paced rest is greater for the heavier loading. The data exhibited a marked nonhomogeneity of variance, with variance and mean magnitudes positively related, so that analyses of variance are not presented. If it is assumed that at each prerest duration the loadings have equal probability of obtaining the longer self-paced rest, the binomial expansion 2 $(.5)^5$ gives a probability of .0625 for either loading's yielding the longer rest at all five durations of prerest cranking.[3]

The grand mean difference between the self-paced rests taken after work against Loads 0 and 4 is about 15 sec. The smallest difference (9.5 sec.) is associated with the 10-sec. duration of prerest cranking, the largest (20.8 sec.) with the 120-sec. duration. Thus, though Fig. 1 suggests little interaction of the rest effects of the two experimental variables, the between-loading difference in self-paced rest follows a trend roughly resembling the within-loading trend.

Some indication that self-paced rest is related to inhibition is evident in the increasing trend of either curve in Fig. 1. Other checks are, of course, available in comparisons of the pre- and postrest starting rates (number of revolutions in the first 10 sec. of a period) of groups with prerest work against Load 0, and in comparisons of the postrest performances of the 10 different groups. The latter checks suggest that Ss did not allow sufficient rest for complete recovery. The average difference between pre- and postrest starting rates of the Load 0 groups is a little greater than one revolution in favor of prerest cranking. The direction of the difference between starting rates in the two periods was consistent for the five groups; the difference is of slight magnitude, however, amounting to less than the difference between Trials 1 and 2 of prerest cranking. The performances of the 10 groups, moreover, tended to vary with both prerest variables: the grand mean of the Load 0 groups was slightly greater than the grand mean of Ss who worked against Load 4 in the prerest period, and, averaged over loadings, performance tended to fall off as duration of prerest cranking increased, though the trend was imperfect. Mean and variance magnitudes were somewhat related, but on any one trial the ratio of the largest to the smallest of the 10 variances was no greater than 3.2. Since the within-trial range of variances seemed acceptable, a two-way analysis of variance was performed for each of the first six postrest trials. Only on the first postrest trial, where differential warm-up effects might serve to inflate between-group differences, were the F tests for the load and prerest-duration main effects ($df = 1$ and 390, and 4 and 390, respectively) significant at the .01 level. By the third trial, neither main effect reached the .05 level, and the interaction mean square, with $df = 4$ and 390, was not significant at the .05 level on any of the six trials. Nonetheless, these suggestions of differential recovery over the rest imply that the self-paced rest is less adequate the greater the level of inhibition, and, therefore, that the obtained differences in mean self-paced rest, within and between loadings, are less than would occur with complete recovery.

[3] The study was repeated with four groups ($N = 160$) at different values of prerest duration. Mean self-paced rest after (*a*) 20 sec. of cranking was 85.9 sec. for Load 0 and 94.5 sec. for Load 4, and (*b*) 90 sec. of cranking was 206.9 sec. for Load 0 and 212.3 sec. for Load 4. The probability that seven differences would be in the same direction is .0156.

Discussion

From Hull's expression of the dissipation of I_R as a decay function of rest duration (**4**, p. 300), it follows that the time required for I_R to decay to some constant value increases as the log of the amount of I_R present at the start of rest. If I_R increases with duration of work (presumably with negative acceleration in a self-paced work task), the rest required for I_R to reach a value close to zero should increase with prerest duration. The trend of mean self-paced rest over the first four prerest values met the above expectation, for either loading. The decrease in rest at the longest prerest duration may reflect sampling error or an inadequacy in the method. The performance data did suggest that *S*'s estimate of the rest required for recovery was less adequate, the greater the presumed amount of inhibition. If the present measure is accepted as related to inhibition, though perhaps imperfectly, the data provide evidence of differential between-loading accrual of I_R. The resting data give valuable support to previous findings, in that self-paced rest is free of the objection that habit bias could account for between-loading differences. A habit bias of a different category may be introduced, of course, if *S* shares *E*'s stereotype of the effect of work loading on the accrual of inhibition.

On the assumption that the difference between self-paced rests is a linear function of the difference between the logs of the amounts of inhibition present at the start of rest, between-loading differences in self-paced rest can also offer indications as to the trend of between-loading differences in I_R as practice progresses. For example, if the between-loading difference in inhibition is quickly stabilized at some constant value, the between-loading difference in rest should have a negative slope with increasing durations of prerest work. If between-loading differences in rest are to increase, between-loading differences in I_R must increase with positive acceleration as prerest work increases. For the present, the obtained differences in self-paced rest may be tentatively accepted as without trend, and as suggesting that approximate proportionality obtains between the amounts of I_R accrued against Loads 0 and 4 in any given duration of work.

Summary

Basic trainee airmen were used to explore the effects of prerest work loading and duration of prerest work on the duration of self-paced rest. Self-paced rest showed some tendency to increase with prerest work, and at each value of prerest work, the heavier of two work loadings yielded the longer rest.

References

1. Bilodeau, E. A. Decrements and recovery from decrements in a simple work task with variation in force requirements at different stages of practice. *J. exp. Psychol.*, 1952, **44**, 96–100.
2. Bilodeau, E. A. The effects of force variations before rest on rate of responding after rest. *USAF Hum. Resour. Res. Cent., Res. Bull.*, 1953, No. 53–26.
3. Bilodeau, I. McD., & Bilodeau, E. A. Some effects of work loading in a repetitive motor task. *J. exp. Psychol.*, 1954, **48**, 455–467.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

(Received November 5, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

THE ROLE OF EXTRAMAZE CUES IN PLACE AND RESPONSE LEARNING[1]

DONALD P. SCHARLOCK

Roosevelt University

Some years ago Blodgett and McCutchan (**2**) noted that one of the basic problems involved in experimental investigations of the spatial learning hypothesis was that of developing a procedure which clearly separates a place disposition from a response disposition. Furthermore, it was their contention that the procedures employed by Tolman, Ritchie, and Kalish (**8, 9**) were not adequate for this purpose. In this regard they remark that "the relatively slow learning shown by the Tolman, Ritchie, and Kalish response group is attributable to external inhibition derived from the fact that spatially opposed responses occurred at the locus of the choice point" (**2**, p. 17). The experiments to be reported in this paper were designed (*a*) to test the hypothesis that the situations employed by Tolman, Ritchie, and Kalish (**8, 9**) for comparing the relative rates of acquisition of place and response learning tended to obscure the actual rate of response acquisition due to inadequate control of the extramaze environment, and (*b*) to examine the role of extramaze cues in the performance of place and response groups trained by the methods described by these investigators.

[1] This article is excerpted from a dissertation submitted to the faculty of the Graduate School of Arts and Sciences at the University of Buffalo in partial fulfillment of the requirements for the Ph.D. degree. The author is greatly indebted to Dr. B. Richard Bugelski for his assistance, guidance, and encouragement throughout the course of this research.

METHOD

Apparatus.—The apparatus employed in the present study has been described in detail in a previous publication (**5**, p. 32). Essentially it consisted of a special muslin "room" constructed to bring *S*s visual environment under *E*'s control and containing an elevated T maze. The overall length of the goal arm was 9 ft. The starting arm was 4 ft. long and perpendicularly incident to the goal arm at its mid-point. It was detachable from the goal arm and could be varied in position depending upon whether *S* was required to start from the north or south side of the room. Food cups were affixed to the ends of the goal arm. Extramaze cues were provided by two 7½-w. lamps located 6 in. beyond the food cups on the longitudinal axis of the goal arm. During the training sessions *S*s were housed in a cage outside of the experimental room.

Subjects.—Ninety-two albino rats, 40 males and 52 females, ranging in age from 90 to 120 days and having no previous experimental history, were selected from the stock of the animal laboratory maintained by the Department of Psychology at the University of Buffalo.

EXPERIMENT I

Method.—All *S*s were placed on a 24-hr. maintenance schedule 48 hr. prior to habituation. Three days of habituation on 2 × 4-in. pine rails set up for that purpose in another room of the laboratory were administered. The *S*s were permitted to run and feed on these rails for approximately 15 min. each day. On Day 4, to examine for position habit, four trials were given to each *S* on the T maze outside of the experimental room, with 90° rotation of the maze after each trial. During these trials food was to be found in both food cups. If *S* exhibited a preference for right turns, i.e., three out of the four trials were right turns, it was required to turn left during the experiment proper, if it was in a Response Group. Similarly, right turns were required of *S*s exhibiting left-turn preferences. The *S*s in the Place Groups either exhibited no particular turn preference, i.e., two right turns and two left turns in the test for position habit, or they were assigned to these groups in such a manner that there were as many right turners as left turners in a given group.

Beginning on Day 5 all *Ss* were given four trials a day by the noncorrection method. Seven successive days of training (Days 5–11) were given, with the starting positions ABBA on the odd days and BAAB on the even days. Intramaze cues were varied by rotating the arms of the T around the choice point between trials. In general, training procedures followed those described by Tolman, Ritchie, and Kalish (9), with the exception of a longer interval between trials, i.e., at least 10 min. for any given *S*. A recent report by Thompson and Thompson (6) indicated the desirability of increasing the interval between trials since massed practice tends to increase reactive inhibitory factors for the response group, hence placing its performance at a disadvantage. Four main experimental groups were employed: (*a*) *1-Lamp Response Group* ($N = 20$). These *Ss* were trained to make a particular turn (right or left depending on position habit) with the lamp at the west end of the T maze always lighted. Hence any *S* in this group always made the same turn regardless of starting position (north or south). On any particular day's training *S* would have to run twice toward the lighted lamp and twice away from the lighted lamp. The *Ss* in this group are confronted with different patterns of visual stimulation depending upon their direction of approach to the choice point. (*b*) *2-Lamp Response Group* ($N = 20$): The *Ss* in this group were trained to make a particular turn (right or left depending on position habit) with the lamps at both ends of the T maze lighted simultaneously. Any *S* always made the same turn regardless of starting position and always ran toward a lighted lamp. For this group the extramaze cues are symmetrically placed with respect to the choice point. Theoretically, these *Ss* should be unable to discriminate one starting position from the other and hence could not acquire a "place disposition." They are always confronted with the same pattern of visual stimulation regardless of their direction of approach to the choice point. (*c*) *1-Lamp Place Group* ($N = 20$): The *Ss* in this group were trained to go to a constant location in the experimental room, half to the east end of the maze and half to the west end, depending upon the position of the lighted lamp. These *Ss* were required to go to the lighted "place" regardless of the turn involved. (*d*) *2-Lamp Place Group* ($N = 20$): The *Ss* in this group were trained to go to a constant location in the experimental room, half to the east end of the maze and half to the west end, regardless of the turn involved. Both lamps were lighted simultaneously. These *Ss* were required to learn the "place" in a situation where no differential directional cues were provided. This group served as a control on the homogeneity of the extramaze environment and on the possibility that a temporal sequence of kinesthetic processes could be learned in this situation.

Results.—Table 1 summarizes the observations made on the performance of the four groups throughout training. The means are plotted in Fig. 1.

An analysis of variance applied to the data indicated that (*a*) the trend in the training session means was not the same for all cue arrangements ($F = 5.63$; $p < .001$), and (*b*) significant variance due to cue arrangement existed ($F = 48.09$; $p < .001$).

TABLE 1

MEAN ERRORS DURING THE SEVEN TRAINING SESSIONS
($N = 20$ each group)

Training Session	Place Groups				Response Groups			
	1-Lamp		2-Lamp		1-Lamp		2-Lamp	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	2.55	1.56	2.20	1.03	2.45	1.20	2.20	1.08
2	2.65	1.15	2.05	.98	1.60	1.68	.80	.60
3	1.75	1.26	2.50	1.12	.90	.26	.85	.27
4	.90	.99	1.95	1.20	.85	.26	.70	.90
5	.95	.97	1.95	.94	.60	.23	.30	.55
6	.60	.92	2.20	.99	.25	.43	.10	.54
7	.25	.62	1.85	.96	.05	.21	.00	.00
Mean	1.38		2.10		.96		.71	

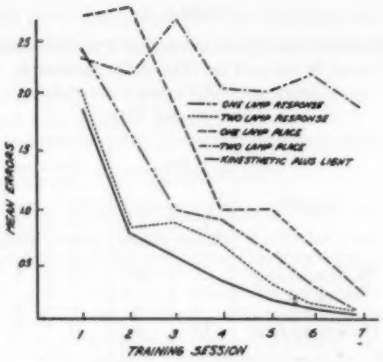

FIG. 1. Learning curves for each of the four groups used in Exp. I and for a fifth group used in Exp. III. The plot is the mean number of errors per training session.

Subsequently, *t* tests of the significance of the difference between pairs of over-all means were made. The error component of the "between-*S*s" variance was used in the computation of *t*, and *t* was evaluated on the basis of 76 *df*. These tests showed all such differences to be significant at better than the .05 level.

Discussion.—The relative performances of the two Response Groups in this experiment can be taken as support for the hypothesis that the Tolman, Ritchie, and Kalish (**8, 9**) situations for comparing the rates of acquisition of place and response learning tended to obscure the actual rate of response learning. The data suggest that in a situation where other extramaze cues are controlled, *S*s will learn a T-maze problem on the basis of kinesthetic cues. Furthermore, such learning will occur more rapidly than that of a group of *S*s trained to take two different paths to the same goal location even in an environment where there are marked extramaze cues. Accordingly, the conclusion of Tolman, Ritchie, and Kalish to the effect that place learning is superior to response learning is not beyond denial.

Ritchie, Aeschliman, and Peirce (**4**) have suggested that neither place nor response dispositions can be acquired in a situation where the animal cannot discriminate one starting position from the other. It might follow then, that neither the 2-Lamp Place Group nor the 2-Lamp Response Group employed in this study were really "place" and "response" groups in their sense of the terms. As a matter of fact, this assumption is partially confirmed by the failure of the 2-Lamp Place Group to learn the problem. However, it is evident that the 2-Lamp Response Group learned something. Can it then be inferred that the categories of "place" and "response" disposition are not exhaustive of the learning possibilities in the T maze?

The same authors have criticized the Blodgett and McCutchan (**1, 2**) studies which suggested that response learning was at least as easy as place learning and in some instances even superior. They point out that the results obtained by Blodgett and McCutchan were related to a number of factors: (*a*) the absence of a definitive directional cue, (*b*) failure to allow *S* to eat on the maze at the food location, and (*c*) the use of the noncorrection method in training. In the present experiment the directional cue provided a marked gradient of illumination throughout the experimental room. The fact that the 1-Lamp Place Group did learn the problem indicates that this cue was sufficient to support the behavior in the absence of reliable kinesthetic cues. The *S*s in this experiment were permitted to eat on the maze at the food location and hence could gain "information" regarding the spatial relationships between this location and other aspects of the environment. Finally, it is difficult to understand the third criticism from the point of view that Tolman, Ritchie, and Kalish (**8, 9**) have reported the superiority of place learning under both the correction and noncorrection methods of training.

EXPERIMENT II

An incidental observation made during the course of training *S*s in Exp. I suggested a further line of

investigation. It was observed that when the distance between the food cup and the lighted lamp was increased, *Ss* in the 1-Lamp Place Group tended to run off the end of the maze. The directional cue was apparently not serving as a mere "sign" of the goal location. Kendler and Gasser (**3**), and Wilcoxon and Waters (**10**) have emphasized the point that place learning is but a special case of discrimination learning. They suggest that it involves the learning of a response (moving toward) certain features of the extramaze environment. The observation that *Ss* tend to run or fall off the maze under the conditions described tends to support their suggestion. It will be recalled that Tolman, Ritchie, and Kalish (**8**) have maintained that the directional cues in their various experimental environments function as "signs" and not as conditioned stimuli. Accordingly, the following procedure was designed in an effort to obtain data which would relate to this hypothesis.

Method.—The *Ss* employed in Exp. I were also used in this experiment. During their training the lighted lamp had always been located 6 in. beyond the food cup. Following the last trial of the seventh training session (Day 11), the lighted lamp was moved to a position 20 in. beyond the food cup, and each *S* was given a single test trial.

Results and discussion.—On the basis of their performance on the test trial *Ss* were assigned to the following three categories: (*a*) those making a successful stop at the food cup, (*b*) those falling into the food cup, and (*c*) those running over the food cup and falling to the floor. Table 2 summarizes the observations.

From these data it may be inferred that *Ss* in the 1-Lamp Place Group are, in fact, running to the lighted lamp rather than to a specific place signified by the lamp. Training, involving differential reinforcement at opposite ends of the goal arm, seems to have established the discrimination of the lighted end from the darker end. The performance of the place *S* depends primarily on the lamp, the turn being suited to its position. The 2-Lamp Place Group failed to show any improvement in performance with increasing amounts of training. This group could not have developed a place disposition since its environment was devoid of discriminable extramaze cues. It was also impossible for this group to have learned a discrimination. The 1-Lamp Response Group continued to perform as usual on the test trial since its training was such as to facilitate the learning of a discrimination based on kinesthetic cues. The *Ss* in this Group found food sometimes at the lighted end of the maze, sometimes at the darker end. Hence a discrimination based on a constant relationship between the lighted lamp and the food could not be learned. The 2-Lamp Response Group exhibited a strong tendency to run off the end of the maze. However, inspection of Table 2 suggests that such tendencies are not as strong in this group as they were in the 1-Lamp Place Group. The 2-Lamp Response Group learns

TABLE 2

CATEGORICAL DISTRIBUTION OF PERFORMANCES OF *Ss* ON A TEST TRIAL FOLLOWING A SHIFT IN THE POSITION OF THE DIRECTIONAL CUE

Category	Place Groups		Response Groups	
	1-Lamp	2-Lamp	1-Lamp	2-Lamp
Successful stop	2	20	19	4
Falling into food cup	4	0	1	8
Falling to floor	14	0	0	8

the problem primarily on the basis of kinesthetic cues, but its performance is also supported by a consistent relationship between the food location and the lighted lamp.

Experiment III

The data discussed above support an interpretation of the performance of the four experimental groups based on discrimination learning. In addition, these findings suggest that a 1-Lamp Response Group, trained in the usual fashion but always running toward a lighted lamp (or away from it), should exhibit the most rapid learning. The performance of such a group would be supported by both kinesthetic and visual cues.

Results and discussion.—Table 3 summarizes the performance of this group. These data are also plotted in Fig. 1. No significant difference exists between this group and the 2-Lamp Response Group. It may be that the additional facilitation of performance brought about by the addition of the light cue is not very great, especially when errors are taken as the measure of performance.

As was the case with the four main experimental groups, following the last trial on the seventh training session all *S*s in this group were given a test trial with the lamp 20 in. from the food cup. On this test trial 8 *S*s fell to the floor, 1 fell into the food cup, 2 made successful stops, and 1 fell off the maze in short-cutting from the starting arm to the goal arm en route to the lamp. These results provide additional support for the hypothesis that the light cue functions as a conditioned stimulus. Although the addition of an extramaze cue to the situation in which *S* can utilize proprioceptive stimuli may not significantly decrease the number of errors, it clearly controls *S*'s "stopping" behavior.

TABLE 3

Mean Errors Made by the Kinesthetic Plus Light–Cue Group during Seven Training Sessions

Training Session	1	2	3	4	5	6	7
Mean	2.10	.78	.58	.30	.17	.08	.00
SD	1.00	.46	.64	.46	.37	.26	.00

Ritchie, Aeschliman, and Peirce (4) have maintained that although the learning of a simple T-maze problem is based, in the early stages, upon the acquisition of place disposition, continued practice produces a kinesthetic habit. The data presented here suggest that even with considerable practice extramaze cues still play an important role.

Summary

Five groups of 20 rats were given four nonconrrection trials for seven successive nights on a T maze. The extramaze environment was controlled. Five different situations involving various combinations of kinesthetic and visual extramaze cues were employed. It was found that in these situations: (*a*) response learning was superior to place learning, (*b*) the performance of a response group trained in an environment devoid of confusing extramaze cues tended to be superior to that of a response group trained in a situation where such cues were present, (*c*) the situation originally designed by Tolman, Ritchie, and Kalish to separate place and response learning tends to obscure the rate of acquisition of kinesthetic habits, and (*d*) place performance and response performance are both based on learned discriminations, the cues supporting the former being primarily visual and those supporting the latter primarily proprioceptive. The relationship of these findings to those reported by other investigators is considered.

References

1. Blodgett, H. C., & McCutchan, K. Place versus response learning in the simple T maze. *J. exp. Psychol.*, 1947, **37**, 412–422.

2. Blodgett, H. C., & McCutchan, K. Relative strength of place and response learning in the T maze. *J. comp. physiol. Psychol.*, 1948, **41**, 17–24.
3. Kendler, H. H., & Gasser, W. P. Variables in spatial learning. I. Number of reinforcements during training. *J. comp. physiol. Psychol.*, 1948, **41**, 178–187.
4. Ritchie, B. F., Aeschliman, B., & Peirce, P. Studies in spatial learning: VIII. Place performance and the acquisition of place dispositions. *J. comp. physiol. Psychol.*, 1950, **43**, 73–85.
5. Scharlock, D. P. The effects of a preextinction procedure on the extinction of place and response performance in a T maze. *J. exp. Psychol.*, 1954, **48**, 31–36.
6. Thompson, M. E., & Thompson, J. P. Reactive inhibition as a factor in maze learning: II. The role of reactive inhibition in studies of place versus response learning. *J. exp. Psychol.*, 1949, **39**, 883–891.
7. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning: I. Orientation and the short-cut. *J. exp. Psychol.*, 1946, **36**, 13–24.
8. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning: II. Place learning versus response learning. *J. exp. Psychol.*, 1946, **36**, 221–229.
9. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning: V. Response versus place learning by the noncorrection method. *J. exp. Psychol.*, 1947, **37**, 285–292.
10. Wilcoxon, H. C., & Waters, R. H. Spatial orientation in the white rat. *J. exp. Psychol.*, 1948, **38**, 412–419.

(Received November 9, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

THE FACILITATING EFFECT OF WORDS ON DISCRIMINATION AND GENERALIZATION [1]

DORIS DIETZE

University of Washington

From an S-R point of view the formation of concepts involves two related processes—discrimination and generalization. When a person learns to respond selectively to similar stimulus situations, he is said to evidence discrimination; and when he learns to respond in the same way to different situations which serve equivalent functions, he is said to evidence generalization. (Generalization used in this sense refers to learned or "secondary generalization," as distinct from primary generalization, which occurs when stimuli evoke similar responses without previous learning.) The facilitating effect of language in the establishment of these two processes may be formulated theoretically as follows: Stimuli produced by responding with highly distinctive names to different situations should tend to increase the difference between stimulus patterns, lessening primary generalization of responses from one of these situations to another. Increased differentiation based upon this process has been called "acquired distinctiveness of cues" (**4**). Stimuli produced by responding to different situations with the same verbal response should become a common element of the stimulus patterns, mediating increased amounts of generalization from one situation to another. This has been called "acquired equivalence of cues" or "secondary generalization" (**4**).

These theoretical statements imply that correct attachment of distinctive names should be easier to learn than correct attachment of very similar sounding names, since there would be less primary generalization during the learning process. With fewer intrusions based on primary generalization, secondary generalization should proceed more rapidly.

In the usual experimental concept formation situation the two processes, discrimination and generalization, probably occur conjointly. However, many studies of concept formation (**9, 10**), by quantifying results in terms of the number of trials or the number of errors before achieving a final criterion, necessarily confound the two. Sometimes they are then inferred separately from introspective or spontaneous comments by the *Ss*. Some descriptions of human *Ss*' concept attainment give the impression that human learning differs in its basic processes from animal learning, by use of vague or poorly defined terms like "active search," "thing-like aspects," or "composite photograph" (**9, 10**).

Investigations of the attainment of concepts have been conducted largely at the adult level. However, in adults established problem-solving habits are applied to new learning situations, and what is observed may reflect not so much the acquisition of new concepts as transfer from previous experiences. When *Ss* are young children, for whom language is a

[1] Adapted from a thesis presented in partial fulfillment of the requirements for the M.S. Degree, Psychology Department, University of Washington, 1953. The writer is greatly indebted to Dr. Allan Katcher for his capable guidance.

relatively new tool, the situation is likely to reveal more about the learning processes. Too often studies of concept development in children have been more concerned with their ability to conceptualize or with their repertory of concepts than with psychological processes. In several of these studies, however, superior performance was correlated with adequate verbalization during the acquisition process (**2, 3, 5, 6, 7, 9**). Additional evidence for the significance of language in discrimination was presented by Pyles (**8**), who found that children demonstrated superior performance in the discrimination of named, rather than of unnamed, nonsense figures.

For the present study a simple concept formation arrangement was designed to examine the part played by words in discrimination and generalization. Of particular interest was the effect of differences in the distinctiveness of names on three stages of learning to name objects: (*a*) initial discrimination, (*b*) subsequent discrimination of related objects, and (*c*) secondary generalization to related objects.

Method

Subjects.—The *S*s were 48 preschool children, ranging in age from 54 to 67 months (mean age: 60.6 months). The mean age difference between older and younger groups was 201 days; between boys and girls, 18 days; and between similar and different names groups, six days. These children were obtained on a volunteer basis from six day nurseries in Seattle, Washington.

Materials.—Sixty nonsense figures, consisting of a circular core with numbers of projecting tips, were cut from masonite. The over-all size of these forms varied with the number of projections, but the core was a standard 1½ in. in diameter, while the projections were ¾ in. wide and 1½ in. long. The figures were varied in four ways: (*a*) the numbers of projections ranged from 1 to 16, (*b*) the spatial arrangement or grouping of projections varied, (*c*) the tips of projections on each figure were all either rounded, pointed, or squared, and (*d*) the figures were painted three slightly different shades of red. Twenty sets of forms were prepared, each set consisting of three forms alike in number and spatial arrangement of projections, but differing in shape of tips and shade of red. The sets were arranged in series, with each set differing from previous sets in number and arrangement of projections, but containing one form with each shape of tips and shade of red. All forms with a particular shape of tips and shade of red were given the same name, regardless of number and spatial arrangement of projections; all forms alike in number and arrangement of tips had different names. Thus shade of red and shape of tips were cues upon which discrimination and secondary generalization were based. To prevent direct comparison of the common details, each set was kept in a separate numbered envelope at all times except when used.

Tiny trinkets of the type dispensed in chewing-gum vending machines were used to keep motivation at a high level.

Experimental design.—A $2 \times 2 \times 2$ analysis of variance design was used. For half the *S*s, figures were called *jod*, *daf*, and *meep*, nonsense names chosen to have different initial and final consonants and different vowel sounds. For the other half the figures were called *beem*, *meem*, and *peem*, nonsense names chosen to be alike except for the aspiration of the initial consonants, which require practically the same lip formation to pronounce. Otherwise the two groups received identical treatment. Two control variables, age and sex, completed the balanced design. These were included because older children are verbally more experienced than younger and might be expected to show more skill in the type of task required by the experiment; and because several studies (**9**) have indicated that, in the age range considered here, girls show superior verbal ability compared to boys.

Procedure.—The *S*s were chosen alternately for each condition from prepared lists upon which they had been assigned randomly to one or the other experimental condition. The selection was done in such a way that an equal number of boys and girls and older and younger children were tested at approximately the same time of day and during each third of the period of weeks during which testing was conducted. This precaution was taken to ensure balance among the groups of any effects dependent on regular nursery school schedule, fatigue, familiarity with *E*, changes in *E*'s attitude, etc. Since *S*s were tested from one day school at a time, this balance throughout the weeks of testing was also desirable to randomize effects of differences in intelligence, training or experience, and socio-economic status that might exist from one school to another.

Testing was done with each *S* alone in a room separated from the regular playroom. The *E* demonstrated the first set of forms, saying, "These are the things we play with." The *E* pointed to each form, pronouncing its assigned name and asking *S* to repeat the names. The *S* was then asked to say the names alone. Care was taken to ensure that he could pronounce and remember all three names without aid from *E*, since, in order to use the names as an aid in discriminating the forms, he had to be able to discriminate the pronunciation of the names. When *S* had mastered the pronunciations, *E* rearranged the forms and said, "As soon as you can name these three blocks without my helping you, I will put one of these little toys (showing the trinkets) into this jar. Now, I have some more of these (indicating and naming the figures), and when you can name the first set we will go on and name the second set. Each time you can name them without my helping you, you can choose a toy, and at the end of the game, when you can tell me all the names without me ever telling you, I will take all the toys out of the jar and make you a bracelet. Do you understand how to play?" The usual response was an enthusiastic attempt to name the figures. The *E* then arranged the figures on the table and asked *S* to identify each form successively. Whenever *S* misnamed one, *E* renamed all three, rearranged them on the table, and again asked *S* to identify the figures, each time in a different succession. The form of the question was varied to maintain interest by such devices as, "Which one is nearest you?," "Put your thumb on the middle one—which one are you touching?" Questions in which *E* pronounced the name before *S*, like "Which one is *daf*?," "Put *jod* in your lap," were alternated with those requiring *S* to say the name first; but in every case *S* pronounced the name while indicating the appropriate figure. Effort was made to structure the situation so that each *S* had equal opportunity to verbalize during learning. Work with Set I continued until *S* correctly indicated all three forms without prompting on three successive trials. At this time he chose a trinket, Set I was put away, and Set II was placed before him. The *E* asked for the names as before, and the "game" continued. When *S* correctly anticipated the names on three consecutive sets, the final learning criterion was reached. This was accomplished in one session averaging about one-half hour. Interest and motivation appeared very high throughout.

Measures.—Records were kept of the number of prompts necessary to learn to criterion each set of three similar but not identical forms. Also recorded was the number of sets (different from the first but related to it by common physical details associated with the same three names) which were required to complete the task. A prompt consisted of renaming all three forms each time an error was made on a single response. One correct trial included three consecutive correct responses, one to each form. The criterion of discrimination learning of a particular set, three consecutive correct trials, consisted of nine consecutive successful responses. The largest number of prompts given on a particular set was 27 for the slowest *S*. The criterion of secondary generalization, correctly naming three

TABLE 1

MEANS AND *SD*'S OF EACH SUBGROUP IN THE ANALYSES OF VARIANCE
(Each figure based on 6 cases)

Measure		Similar Names					Different Names				
		Boys		Girls		Total	Boys		Girls		Total
		Older	Younger	Older	Younger		Older	Younger	Older	Younger	
Set I	*M*	2.86	3.08	2.96	2.05	2.74	2.81	2.28	2.17	1.82	2.27
	SD	.39	1.15	.96	.52	.91	.88	.35	.54	.41	.68
Prompts per set	*M*	1.59	1.90	1.93	1.75	1.79	1.71	1.48	1.20	1.34	1.43
	SD	.23	.43	.52	.44	.44	.32	.42	.23	.36	.39
Number of sets	*M*	2.50	2.50	3.04	2.77	2.71	1.89	1.44	1.58	1.86	1.69
	SD	.64	.62	.63	.38	.62	.66	.52	.41	.60	.59
Total prompts	*M*	3.60	4.02	5.17	4.21	4.25	2.42	1.94	1.68	2.28	2.08
	SD	1.53	.70	1.33	1.45	1.42*	.56	.87	.43	.90	.78*

* Variances significantly different at the .01 level.

sets three times each, involved 30 consecutive correct responses. The largest total number of prompts given was 73. Thus, the criteria of discrimination and generalization were sufficiently high to eliminate practically all possibility of chance success (1).

Four measures were considered in the analysis of the data: (*a*) number of prompts given during Set I was the measure of initial discrimination; (*b*) number of *prompts per set*, excluding Set I, was the measure of discrimination on sets after the initial discrimination; (*c*) *number of sets* completed after Set I and before reaching the criterion was the measure of secondary generalization; (*d*) *total* number of prompts was used as a general measure of efficiency of concept formation, involving both discrimination and generalization.

Because the ranges of scores were limited at the lower end, the distributions were highly skewed. To normalize them a square-root transformation was made, which produced the desired effect, with one exception on total prompts (see Table 1).

Four analyses of variance were performed, using data of each of the four measures of learning.

Results

Table 1 shows the means and *SD*'s of the eight subgroups on each of the four measures.

The largest difference between the similar-and different-names groups was on total prompts, the score which incorporated the other three measures. This *F* ratio was 43.05. (The magnitude of this ratio, however, may be questioned, since the variance of the similar-names group was significantly larger than that of the different-names group.) For prompts per set, the measure of subsequent discrimination, *F* was 8.89, and for number of sets, the measure of secondary generalization, *F* was 31.98. All three of these values exceed 7.31, the value of *F* required with 1 and 40 *df* for the .01 confidence point. The initial discrimination, Set I, yielded an *F* of 4.33, which is larger than the 4.08 required for the .05 confidence point. In each case the different-names group had the smaller mean.

Intercorrelations between measures, except for total prompts, were low: $-.10$ to .42. (For 22 *df* *r*'s higher than $\pm .40$ are different from zero at the .05 level). The correlation between prompts per set and number of sets was $-.10$ for the similar-names group and .31 for the different-names group.

On Set I prompts there was a difference between boys and girls significant above the .05 point ($F = 5.15$) showing superior performance by girls. With this exception, the *F* ratios for the control variables and the interactions were not significant.

Discussion

Kind of Names

Discrimination.—Results of the analyses of variance show clearly that the children learned to name similar appearing, nonconventional forms more easily when distinctive names were used than when similar names were used. Furthermore, they learned to apply these names to new forms more readily when the names were distinctive. Interpretation of these results, assuming the same principles that underlie simpler conditioning and discrimination problems, suggests two related phenomena—stimulus generalization and response generalization.

Before beginning the discrimination task, all *S*s were familiarized with the nonsense names to be used to the point where they could recite all three without prompting. (The six words used in the study were considered equivalent in that each had a constant consonant-vowel-consonant construction. Although the different-names group actually had a larger number of sound combinations to learn, this initial phase was so brief for both groups that any differences in learning difficulty were assumed to be inconsequential.) In the discrimination situation the child was actually confronted with two different types of tasks: after receiving three similar ap-

pearing forms he had to (*a*) choose the correct form upon hearing each of the names, and, (*b*) say the correct one of the names as *E* pointed to each form.

In the first type of task both groups were required to make the same overt responses, such as pointing or putting the form in the lap. Only if these responses were made to stimulus configurations which included more differentiating cues would the different-names group be expected to appropriately perform more readily. Since the same figures were used for each group, these differentiating cues must have existed in the names used.

In the second type of task, when *E* pointed to a specific form, *S* had three responses available. If he said the correct name, the combination of the visual stimulus of the form and the auditory and proprioceptive stimuli arising from pronouncing its name would be reinforced. Nonreinforcement of incorrect responses would tend to cause combinations of this visual form and response-produced cues from other names to lose relative strength. In the early stages of learning errors were frequent. In the different-names group, where available responses were (relatively) dissimilar, response generalization would be less likely than primary stimulus generalization among similar visual forms. In the similar-names group, however, both visual stimuli and available responses were similar; hence errors may be accounted for by response generalization. But in what sense are these responses similar? Similar stimuli, auditory and proprioceptive, are produced and received by pronouncing them. Thus again primary stimulus generalization can account for the less adequate performance of the similar-names group. In brief, stimuli produced by responding with the correct name became part of the visual-proprioceptive stimulus configuration, decreasing primary generalization among the stimulus patterns. This effect was more rapid when differentiating cues supplied by the names were distinctive.

In the very early stages of learning, conflicting responses (like "jaf" or "mod") somewhere between two of the prelearned responses, occasionally occurred in the different-names group. In the similar-names group, where such exchange of initial consonants would merely produce another of the available names, an occasional substitute, such as "leem" or "keem" was produced. This type of response may have appeared as evidence of close balance between the strength of available responses, and it may have appeared as response generalization.

Secondary generalization.—A theoretical account of the results of the secondary generalization task poses quite a problem in the present framework. With fewer intrusions based upon primary generalization, secondary generalization should proceed more rapidly. However, nothing in the present data indicates that such intrusions were more frequent in the similar-names group. In fact, the absence of significant correlation between prompts per set and number of sets may be interpreted as evidence that performance in the secondary generalization task was not associated with discrimination. Additional research may provide an answer to this provocative situation.

Control Variables

Since previous studies have indicated age differences in verbal learning speed, age was included in the present design as a control variable. It is not surprising, however, that no significant differences attributable to age appeared. First, there was less than a year's difference between the older and younger groups. Second, when children are beginning to acquire language skills, development is so rapid that small age differences may be expected to produce large differences in relative verbal skill. However, *S*s of this experiment were past the age period where verbal development is relatively most rapid.

The one significant sex difference, on Set I (initial discrimination), conforms to the results of previous investigations which show that girls in this age range

have a slight edge over boys in verbal learning. That this superiority was not maintained over the other measures suggests that girls may be superior in the initial rote-learning stage of acquisition, but that there is very little difference in generalizing ability once the names are learned.

Summary

It was hypothesized that complex discriminations between similar appearing forms are easier if names applied to forms are distinctive than if names are very similar. The study used 48 four- and five-year-old children as *Ss* in learning a simple concept formation task. For the same objects one group learned similar sounding names and the other group learned different sounding names. Four measures of learning were employed to reveal initial discrimination, subsequent discrimination of related objects, secondary generalization to related objects, and general efficiency of concept formation.

Statistical results indicate that on each measure the different-names group learned faster than the similar-names group. For discrimination tasks these results were ascribed to primary stimulus generalization.

References

1. Grant, D. A. Additional tables of the probability of "runs" of correct responses in learning and problem solving. *Psychol. Bull.*, 1947, **44**, 276–279.
2. Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
3. Long, L., & Welch, L. The influence of abstractness on reasoning ability. *J. Psychol.*, 1942, **13**, 41–59.
4. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 173–176.
5. Munn, N. L. *Psychological development.* Boston: Houghton Mifflin, 1938.
6. Munn, N. L. Learning in children. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946.
7. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univ. Press, 1953.
8. Pyles, M. K. Verbalization as a factor in learning. *Child Develpm.*, 1932, **3**, 108–113.
9. Vinacke, W. E. *The psychology of thinking.* New York: McGraw-Hill, 1952.
10. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received November 13, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

LEARNING AS A FUNCTION OF CONTEXTS DIFFERENTIATED THROUGH ANTECEDENT VALUE EXPERIENCE [1]

EUGENE GLOYE [2]

University of Nebraska

This study represents an attempt to clarify some issues associated with the problem of the influence of affective value on learning. In previous studies the affective character of learned materials (usually words) has generally been determined through judgments of "pleasant" (P), "unpleasant" (U), or "indifferent" (I) on the part of *S*s or by *E* himself (**2**). In recognition of the questionable reliability and validity of such procedures, the present study formulates a definition of "affective value" through operations which provide rewarding and mildly punishing associations for certain stimuli under controlled conditions. The "building-in" of values in this manner rests on the reasonable assumption that affective characteristics of stimuli arise through their associations with differentially reinforcing experiences. It must be made clear, however, that the introduction of value differentiating experiences in the experimental situation does not in itself allow the labeling of various stimuli P, U, or I, since these qualities seem to have no status except as judgments (**3**). It is simply assumed that neutral stimuli associated with positive value experiences (reward) will come to have different affective values from neutral stimuli associated with negative value experience (punishment).

The experimental procedure is directed at developing evidence with regard to the general hypothesis that learning of neutral materials is influenced by the values of the contexts in which such materials appear during learning, even though the contexts are not essential to the learning task. More specifically, it was predicted that, on the average, the learning of three-digit numbers appearing in context figures which had neutral value associations would be different from the learning of three-digit numbers appearing in context figures which had positive or negative value associations although the context figures were not part of the required response term in the learning procedure.

METHOD

Subjects.—Forty-eight *S*s were employed in the study: 24 men and 24 women. All were volunteers drawn from introductory psychology classes at the University of Nebraska.

Value development situation.—Various geometric outline figures were differentially reinforced through their use in a card game in which all *S*s participated. A set of cards consisted of 10 such figures printed on playing-card stock with the following values equally often assigned: + 5, + 1, 0, − 1, − 5. Different value-figure relationships were introduced through the use of two sets of cards. In addition, all positive figures appeared in green, all negative figures appeared in red, while neutral figures appeared in black.

The card game was such that *S* was given the opportunity to trade cards with *E* in order to get the highest possible positive score. The

[1] This article reports one phase of a dissertation submitted to the Graduate College of the University of Nebraska in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express his gratitude to Katherine E. Baker for her aid and advice with respect to this project.

[2] Now at the College of Wooster.

value designations for the geometric figures were contained on a poster which was before *S* at all times. It was indicated that the game could be played most efficiently if the figure values were memorized.

The area of investigation was posed to *S*s as concerning "guessing ability" or "intuition." If *S*'s guesses resulted in his obtaining a positive score (group of cards with a positive total), he was paid the amount of his score in pennies. If, however, he finished a trial with a negative score, he was required to pay *E* the amount in pennies. Each *S* participated in 20 trials of the game and *S*s were allowed to keep their winnings after refunding the 25 pennies they were given at the start.

Besides monetary gain, another incentive was used in attempt to strengthen the value differentiation of the figures. Through a pseudo-explanation of a previous study *S*s were led to believe that members of the opposite sex had done better in the game than members of their own sex.

Learning situation.—The *S*s were required to learn three-digit numbers printed inside the previously used geometrical outline figures, the figures now representing "value contexts." A paired-associate method was employed with nonsense syllables as stimuli and numbers as responses. No overt response was demanded for the context figures themselves, though they appeared with the numbers at all times. Thus the figures were in no way essential to the learning task. It was assumed that the numbers learned were neutral because of their random selection. The digits 5, 1, and 0 were omitted from the numbers, however, since they had previously been associated with the figures in the card game.

It was possible to maintain color-figure relationships consistent with those a particular *S* had seen in the card game and at the same time control for associations between particular numbers and particular figures by employing four sets of stimulus cards. The several sets allowed the figure-number relationships to be counterbalanced. A schematic description of the essential features of the four different sets of cards will be found in Table 1.

The nonsense syllables were printed on still another set of 10 cards. The syllables were selected from a list with zero association value given by Hilgard (1). The random initial pairing of syllables and figure-number cards was obtained by simply shuffling the syllable cards for each new *S*. Thus the effect of incidental associations between numbers and syllables was counterbalanced.

An exposure apparatus was employed which allowed the pairs of cards to be presented in a different random order with each new trial so that serial-position effects could be controlled. The apparatus operated in such a way that the syllable associate for each figure-number card was exposed for 2 sec. Following this there was a 2-sec. exposure of both the syllable card and the figure-number card. Between each presentation there was a pause of 2 sec. while the cards were changed manually and there was a 1-min. period between each trial.

TABLE 1

FEATURES OF SETS OF CARDS USED TO PRESENT NUMBERS WITHIN CONTEXTS OF VARIOUS VALUES, COLORS, AND FORMS DURING LEARNING

Value of Figure	Color of Figure	Sets of Cards					
		AI and AII			BI and BII		
		Type of Lines in Figure	Numbers*		Type of Lines in Figure	Numbers*	
			I	II		I	II
5	green	straight	479	694	curved	479	694
5	green	straight	836	827	curved	836	827
1	green	straight	467	648	curved	467	648
1	green	straight	328	926	curved	328	926
0	black	comb.	396	274	comb.	396	274
0	black	comb.	274	396	comb.	274	396
1	red	curved	926	328	straight	926	328
1	red	curved	648	467	straight	648	467
5	red	curved	827	836	straight	827	836
5	red	curved	694	479	straight	694	479

* The numbers are the items learned. They appeared inside the figures.

After a sample trial, *S* was instructed to attempt to anticipate the number which would appear upon seeing its syllable associate. Trials were continued until each number had been correctly anticipated on two successive occasions. When a particular number was learned to this criterion, it was removed from the group. Thus as trials continued, there were fewer and fewer cards presented on each trial.

Results

A graphic presentation of the mean number of trials required by *Ss* to learn the numbers as a function of the values of the context figures in which they appeared during learning will be found in Fig. 1. It may be seen that numbers appearing in neutral context figures (designated as "0") were learned most quickly, on the average. Further, an increase in value intensity, without regard for direction (plus or minus) brings about decreasingly efficient learning.

The number of trials required to learn the numbers in each of the value categories was subjected to analysis of variance. Three factors were considered: (*a*) differences between value categories with regard to learning efficiency, (*b*) differences between *Ss* with regard to effects of value contexts on learning efficiency, and (*c*) interaction among these variables. Consequently, a 5 (context values) by 48 (*Ss*) classification was employed, with two scores in each cell (one for each of the two representations of each of the five value categories).

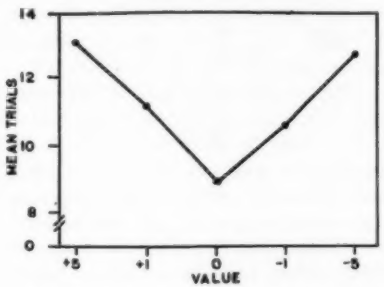

Fig. 1. The mean number of trials required by *Ss* to learn neutral items (i.e., numbers) as a function of the value of the context figures in which they appear during learning.

Interaction yields an *F* of 1.30 which is significant at the .05 level of confidence; consequently, interaction was employed as an error term to determine *F*'s for rows and columns. The *F* for context values is 11.20, a figure highly significant (1% level is 3.41, with 4 and 200 *df*). The *F* for *Ss* is slightly below the .05 level of confidence (1.37, with 50 and 200 *df*). Thus the effect of context value upon learning is demonstrated to be a statistically significant relationship, while there is also a suggestion that *Ss* may react differently to the experimental variable of context value.

Discussion

The results support the hypothesis that value contexts surrounding learning materials will influence learning of such materials. Observed differential efficiency with which neutral materials were learned is clearly dependent upon geometrical context figures appearing with neutral materials since other possible determinants of differential efficiency were carefully controlled.

Within the experimental conditions of the present study it is clear that valued contexts, in contrast to neutral contexts, interfered with learning in the sense that it lowered learning efficiency. The quality of the values (positive and negative) was irrelevant to their influence, while the intensity of the values did bear a systematic relationship to learning efficiency; the higher values apparently produced more powerful value influence.

As a consequence of the results reported here, affective values may be interpreted to be distracting elements over and above the possible distraction provided by the presence of the geometrical figures themselves. While learning performance here is clearly

related to the amount of value given each of the context figures and the interfering properties of the figures seem to mount with increasing value, it is advisable, nevertheless, to maintain some reservation in concluding that the higher the value, the more the learning will be inhibited. Only a small range of values was sampled here and it is possible that more intense values or values based upon ego-involving conditions might produce effects different from those observed. Two other types of relationships might be anticipated: (*a*) various studies suggest that value acts as an emphasizing factor, and (*b*) other results suggest that one might expect positive values to enhance learning while negative values inhibit it in accordance with the need-reduction principle. Neither the emphasis idea nor the need-reduction principle can account for the results presented here, in any case.

Summary

An experiment was conducted to develop evidence related to the hypothesis that learning of neutral materials appearing in value contexts is different from learning of materials appearing in neutral contexts. Values were "built in" for geometrical outline figures through association of rewarding and mildly punishing experiences, and neutral items presented within these context figures were learned by the subjects. Factors such as initial values of the figures, serial position effects, association values of numbers used as neutral items, etc., were counterbalanced so that performance variations could be related to the value differentiation of the context figures.

Results support the hypothesis that value differences influenced learning. Items appearing in neutral contexts were learned more quickly, on the average, than items appearing in value contexts, irrespective of the quality of their values (positive or negative). Evidence supports the principle of value interference rather than either the need-reduction principle or the emphasis principle.

References

1. Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
2. McGeoch, J. A., & Irion, A. L. *Psychology of human learning*. (2nd Ed.) New York: Longmans, Green, 1952.
3. Peters, H. N. The judgmental theory of pleasantness and unpleasantness. *Psychol. Rev.* 1935, **42**, 345–386.

(Received November 10, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

THE EFFECT OF CHANGE IN DRIVE LEVEL ON HABIT REVERSAL[1]

ALEXANDER M. BUCHWALD AND HARRY G. YAMAGUCHI

Indiana University

A series of recently published studies has been concerned with differences in rate of learning between groups of *Ss* selected on the basis of scores on the Taylor Scale of Manifest Anxiety (**7**). Studies involving the conditioned eye blink (**4, 5, 6**) have found high-anxious *Ss* outperforming low-anxious *Ss*, but in studies involving tasks such as learning a verbal maze (**8**) or a stylus maze (**1**) low-anxious *Ss* have tended to reach learning criteria prior to high-anxious *Ss*.

These results have been said (**1, 3, 8**) to be predictable from Hull's theory on the assumption that scores on the anxiety scale are indicative of *S*'s drive level. It is stated that the effects of increases in drive level will depend on the relative habit strengths of the various competing responses.

"If the habit strength of the correct response should be relatively weak, an increase in drive strength should further increase the strength of the incorrect tendencies relative to the correct tendency, resulting in impaired performance" (**1**, p. 120).

To the best of our knowledge this proposition has been tested only in experiments in which *Ss* have been selected on the basis of their Taylor scale scores. It has not been tested in experiments where drive has been established by a deprivation procedure. In the present experiment drive levels based upon water deprivation were used to test the hypothesis that increasing *S*'s drive level will lead to impaired performance when the reinforced response is not initially the strongest in a hierarchy of responses.

[1] The writers thank their colleagues, J. A. Dinsmoor and I. J. Saltzman, for their suggestions and criticisms of an earlier draft of this paper.

METHOD

Subjects.—The *Ss* were 36 experimentally naive female hooded rats bred in the Indiana University laboratory. They were 90 to 135 days old at the beginning of the experiment.

Apparatus.—The apparatus consisted of a single unit T maze with a stem 18½ in. long and arms 18 in. long. All alleys were 3¾ in. wide and 4 in. high and consisted of wooden walls, hardware-cloth flooring, and a hinged cover of Plexiglas. The maze rested on legs which supported it 2 in. above a table top. With the exception of the floor and the cover, it was painted flat black throughout. The starting box was 9 in. long and the goal boxes 9 ¼ in. long. Each goal box had a ¼-in. hole in the wall nearest the starting box, 2½ in. above the floor and 2½ in. from the end of each compartment. A piece of glass tube connected by a section of rubber tube to a water bottle mounted on the outside of the end of the goal box was inserted in this hole with the tip of the glass tube projecting ¼ in. into a goal box. Each goal box had such a water unit. One and one-half inches toward the choice point from each water-tube hole, a baffle, 4 in. high and 1½ in. wide, was affixed to the wall nearest the starting box with the width of the baffle at right angles to the wall. No part of the water units was visible from the stem or choice point. There were four doors in the maze: a guillotine door for the starting box, another door of the same type in the stem at the choice point, and two swinging doors with a vertical sweep, one for each goal box.

Procedure.—The *Ss* were randomly assigned to four groups: high-high, high-low, low-low, and low-high, "low" and "high" referring to low drive level and high drive level, respectively. The high drive condition was effected by depriving *Ss* of water for 20½ hr., the low condition by depriving *Ss* of water for 1½ hr. The *Ss* were habituated and run in squads of four *Ss*, one *S*

from each group. A new squad was started every five days. The *S*s in each squad were run at the same time of the day but were given their daily home-cage watering at different times of the day. Food was always available in the home cages but water was available for only 1½ hours per day. Whenever possible, two *S*s on identical watering schedules were housed in a single cage, but when this was not possible, *S*s were housed singly.

Preliminary habituation consisted of several overlapping phases. The first phase consisted of 10 daily 10-min. periods of handling. On Day 4 of handling, habituation to the maintenance schedule was begun and continued for 10 days. On the last three days of maintenance habituation, *S*s were also habituated to the apparatus.

Apparatus habituation was done in six 5-min. periods, two periods per day, approximately ½ hr. apart. It consisted of two main phases: goal-box habituation and stem habituation. First, *S*s were placed in the goal-box runway with the choice-point door closed and were given water reinforcement for running from one goal box to the other. On the first habituation trial for each *S* no attempt was made to control the location (right or left goal box) of reinforcement. On all subsequent habituation trials the location of reinforcement was alternated. This was done by retracting the water tube from one goal box and inserting the other tube into the other goal box. The *S*s were allowed 10 sec. drinking time for each reinforcement throughout the experiment. Each *S* was given from 10 to 24 habituation reinforcements, inclusively. On the first day of apparatus habituation the goal-box doors were open at all times. On the second and third days, however, they were closed after *S* entered a goal box and started to drink. During stem habituation *S*s were placed in the stem with the choice-point door closed and were allowed to explore. Those *S*s attaining the maximum number of habituation reinforcements before the end of the fourth period of apparatus habituation were put into the stem immediately following the last reinforcement and were replaced in the stem for all remaining habituation periods. All other *S*s were put into the stem during the fifth period and were put into the goal-box runway during the sixth period. Those *S*s attaining the maximum number of reinforcements during the latter period were removed from the runway and put into the stem for the rest of the period. One *S* had not been given the minimum number of reinforcements by the end of the sixth period. For this *S* the sixth period was extended until the minimum was attained.

Original learning trials were started on the day following apparatus habituation. Four trials per day were given until a criterion of at least 10 correct runs in three consecutive days was attained. A choice was recorded as incorrect if *S* had all four feet in the incorrect alley. Correction was allowed on all trials. The intertrial separation was approximately 15 min. On all trials the running procedure was as follows: both goal-box doors were opened and a water tube was inserted in the appropriate goal box; *S* was put into the starting box; the starting-box and choice-point doors were raised simultaneously to start the trial; the starting-box door was lowered when *S* was clear of it and the choice-point door was lowered when *S* had all four feet in either one of the goal-box runways; the goal-box door was closed after *S* had commenced drinking; after 10 sec. of drinking, *S* was removed from the goal box. The *S*s were randomly assigned to subgroups which were reinforced either on the right or left side, this assignment being made with the restriction that in each group four *S*s were reinforced on one side and five *S*s on the other.

The *S*s were not run on the day following the last day of original learning since the deprivation interval for half of the *S*s was changed. This arrangement made it possible for all *S*s to have 1½ hr. of watering preceded by 22½ hrs. of water deprivation prior to the first day of reversal learning.

Reversal-learning trial procedures were identical with those used during original learning with one exception. Each *S* was reinforced for running to the goal box in which it previously had not been reinforced. Reversal-learning trials were continued for each *S* until the same criterion as that used for original learning was met.

Results and Discussion

Table 1 presents the days to criterion for both original learning and reversal learning for all groups of *S*s.

Because of marked skewness and lack of homogeneity of variance present in the data, nonparametric statistics were used for all significance tests. To test for significance of differences in reversal learning, Wilcoxon's test for paired replicates (9) was used. A pair of *S*s consisted of the two *S*s in each squad given original learning at the same drive level.

In reversal learning, each low-high *S* took fewer days to criterion than

TABLE 1

DAYS TO CRITERION FOR ORIGINAL LEARNING AND REVERSAL LEARNING

Group	Days to Criterion					
	Original Learning			Reversal Learning		
	Median	Mean	*SD*	Median	Mean	*SD*
Low-low	10.00	12.33	5.82	11.00	10.44	3.66
Low-high	17.00	22.44	17.06	4.00	4.33	2.45
High-high	4.00	8.11	9.30	5.00	5.22	1.40
High-low	5.00	6.00	4.45	6.00	8.78	4.13

its low-low running mate. This result is significant at the .01 level. Seven of the nine high-high *Ss* reached the criterion more rapidly than their high-low running mates. This result barely misses significance at the .05 level. Thus it would appear that (at least under the conditions of our experiment) a high drive level leads to more rapid attaining of a learning criterion than does a lower drive level when the response being learned has, at the start of learning, a low habit strength relative to that of a competing response.

To check on possible ability differences between *Ss*, the same test was run on days to criterion in original learning. In neither case (high-high vs. high-low; low-low vs. low-high), do these results approach the .05 level of significance.

The use of days to criterion as the response measure in the present experiment may serve to obscure an initially greater resistance to abandoning the originally learned response on the part of the groups that had reversal learning under high drive. An analysis of the data, however, does not support this hypothesis. When *Ss* are divided according to whether they fall above or below the median of the distribution of number of reversal responses made on the last three trials of the first reversal day, we find a slight but insignificant tendency ($\chi^2 < 1$, $df = 1$) for *Ss* undergoing reversal under high drive to make more new responses than do *Ss* undergoing reversal under low drive.

Since the design of the present experiment is similar to that of experiments investigating the effects of drive differences during acquisition and extinction, it may be of some value to examine our data for matters other than that which is the primary concern in this study.

Original learning.—When a comparison of days to criterion in original learning is made between low-low and low-high (low-original) *Ss* and high-low and high-high (high-original) *Ss*, we find the high-original *Ss* reaching criterion more rapidly ($P < .01$ using Wilcoxon's test (9) for unpaired replicates) than do the low-original *Ss*. This would seem to indicate that for situations where the habit strengths of two competing responses are initially approximately equal on the average, a higher drive level leads to more rapid learning. This must be stated with caution, however, since the high-original *Ss* had more drinking exposures in apparatus habituation than the low-original *Ss* ($P < .01$; test for unpaired replicates), and this conceivably could have affected their performance in original learning.

Effect of original learning drive on reversal learning.—A comparison of the number of days to reversal criterion for *Ss* given original learning on high drive versus those given original learning on low drive was made by means of the Wilcoxon test for unpaired replicates. The result of this test finds the distributions not significantly different at the .05 level of confidence. As a test of whether the performance of the *Ss* on the first day of reversal learning is influenced by the drive level which prevailed during original learning, the distribution of number of new responses on

the last three trials of the first reversal day was again divided at the median, and the number of *Ss* given original training under the differing drive levels falling above and below the median was counted. There is some tendency for *Ss* given original learning under high drive to make more original learning responses, but this is not significant ($\chi^2 < 2$, $df = 1$).

Effects of changes in drive level.—Comparison of days to criterion in reversal learning reveals that low-low and high-high (unchanged-drive) *Ss* reach criterion more slowly than do high-low and low-high (changed-drive) *Ss*, respectively. In neither case does this difference reach the .05 level of significance (Wilcoxon's test for unpaired replicates) although for high-high vs. low-high it is significant at the .10 level. The test of number of *Ss* falling above and below the median number of reversal responses on the last three trials of the first day of reversal learning shows that the changed-drive *Ss* make more new responses than do the unchanged-drive *Ss* ($\chi^2 = 4.05$, $df = 1$, $P < .05$). This finding is similar to that of an earlier study (**10**) in which a decrement in habit strength resulted from a change in drive level as would be predicted by Hull's theory (**2**).

The primary purpose of this experiment was to present a situation where *S* had to learn a response in a situation where a strong competing response was present. It is, in a sense, an analogue of the Taylor-Spence verbal maze-learning experiment (**8**) and the Farber-Spence stylus maze-learning experiment (**1**). However, both of those experiments found low-anxious (presumably low-drive) *Ss* reaching criterion prior to high-anxious (presumably high-drive) *Ss*.

The chief procedural differences between our experiment and these earlier experiments would seem to be (*a*) a species difference in the *Ss* used, (*b*) different learning histories of the competing responses, and (*c*) differences between the drive variables used.

Of these three, the third alternative, namely, that differences in experimental results are due to the differing natures of the drive variable and/or to the use of selection as opposed to manipulation procedures, would seem to be of primary theoretical importance. If this line of explanation is confirmed by further research, it would seem to suggest the necessity of reinterpretation of the theoretical significance of Taylor scale scores.

Summary

Using a correction procedure, four groups of rats were trained to run a T maze to a criterion of 10 out of 12 correct responses and were then trained to reverse the direction of response. The *Ss* ran under either 1½ or 20½ hr. of water deprivation. At the start of reversal learning the deprivation interval was shifted for two groups but left unchanged for the other two groups.

The results indicate that reversal learning is more rapid under high drive. The discrepancy between this result and results obtained in studies where drive level is defined by scores on the Taylor anxiety scale and the theoretical implications of this discrepancy are discussed.

References

1. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
2. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
3. Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *The Kentucky Symposium: Learning theory, personality theory, and clinical research.* New York: Wiley, 1954.
4. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
5. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
6. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
7. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
8. Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952. **44**, 61–64.
9. Wilcoxon, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Co., 1949.
10. Yamaguchi, H. G. Gradients of drive stimulus (S_D) intensity generalization. *J. exp. Psychol.*, 1952, **43**, 298–304.

(Received November 22, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

WISHING WITH DICE [1]

R. A. McCONNELL, R. J. SNOWDON, AND K. F. POWELL
Department of Physics, University of Pittsburgh

Since 1943 a number of experimental papers have appeared dealing with an effect called "psychokinesis." In a typical experiment, one or more gaming dice were thrown, mechanically or otherwise, while *S* wished the dice to come to rest with certain faces or combinations of faces uppermost. Under some conditions, with certain *S*s, the dice favored the desired faces to a statistically significant degree. The effect was found to persist after the introduction of safeguards, including the use of a motor-driven cage to avoid direct human contact with the dice.

In the preparation of the early experimental data for publication, an unusual distribution of scores was noticed. It was found that *S*'s scoring rate tended to drop off quickly after the beginning of an experimental session, but was raised by any short rest period or interruption during the session. Such scoring declines have been a frequent feature of subsequent papers.

Thus, the evidence for psychokinesis rests upon two statistical effects. These are the total deviation from chance expectation of wished-for die faces, and the occurrence of extra-chance declines in the scoring rate for those faces. From an operational point of view, one might say that the psychokinetic effect *is* those two effects when they appear without a known physical mechanism to explain their occurrence.

Decline and deviation effects are usually associated in the same experiment, but this is not always the case (**2, 4**). Originally, it was supposed that the relaxation obtained while stopping to tally up the score, or when beginning a new data page, was sufficient to rehabilitate *S*'s waning power. It is now known that scoring declines cannot be regarded as a simple fatigue effect (**7**).

The present experiment was undertaken as a search for psychokinesis under somewhat more rigidly controlled conditions than those prevailing in any past experiment, and with the hope that if an effect were found, something new about its nature might be learned. This is the first experiment combining motor-driven machine throwing with photographic recording of the dice.

Among the errors of method and extraneous effects believed to have been successfully excluded in this experiment are the following: record errors, the loss or selection of data, selection of the experiment, retroactive choice of target, *ad hoc* selection

[1] The participation of K. F. Powell in this experiment was made possible by a $1000 grant from the William McDougall Research Fund of Duke University. The die-throwing machine was designed and built by Mr. W. B. Scherer, and upon completion was loaned to the University of Pittsburgh by Dr. J. B. Rhine of Duke University.

The findings of this paper were submitted for criticism to the following members of the faculty of the University of Pittsburgh: Dr. A. Ciocco, head of the Department of Biostatistics; Dr. D. Halliday, head of the Department of Physics; Dr. M. A. Lauffer, head of the Department of Biophysics; and Dr. R. A. Patton, head of the Department of Psychology. Professor Ciocco was assisted in his examination of this paper by Professors P. Densen and D. Horvitz of his staff. Professors Halliday, Lauffer, and Patton also served as subjects during the course of this experiment. The helpful suggestions of these men are gratefully acknowledged.

of the method of analysis, "optional stopping," dice bias, and any hypothetical mechanical "starting effect." Some of these will be discussed in this paper. Others of a purely mechanical or procedural nature have been given auxiliary publication.[2]

Another feature of the present experiment is the use of previously untested, volunteer *S*s. However, in this connection it should be noted that the procedure was one in which *E* was present and aware of the target number. Thus, the word "subject" is used provisionally or in a nominal sense.

Procedure

This experiment includes approximately 170,000 die throws by 393 *S*s over a period of 18 months. The dice were thrown two at a time. The first third of the die throws by each *S* were made from a rough-lined cup into a felt-lined tray. The last two-thirds were thrown by a totally enclosing, plastic cage, 90 cm. long. The cage was turned by a motor about a horizontal axis perpendicular to its length at a rate of 7.5 half-turns per minute. The transparent walls of the cage were studded with soft rubber obstacles to randomize the motion of the falling dice. One pair of dice was used throughout for the cup throwing, and a second pair for the machine throwing.

Volunteer *S*s were recruited from classrooms and among the friends of the *E*s. They were asked to make an appointment to throw dice at an hour of their own choosing in the following two weeks. One *S* and one *E* only were present at every test session. The sessions lasted about 40 min. Each *S* participated only once.

During a preliminary discussion period intended to establish an atmosphere of social informality and relaxation, *S* was asked to accept a target number from one to six from *E*'s list or to choose one in the same range according to his own preference. The *S*'s task was to wish for the falling dice to come to rest with the uppermost faces corresponding individually to the chosen target number. A hit (i.e., a target face) was scored for every die face which agreed with that number (nominal *P* value = $\frac{1}{6}$). The target number and its manner of choice were recorded in pencil along with *S*'s address, etc., on the data page before the throwing began. The same target number was repeatedly recorded throughout the test session, six times on every data page—primarily to keep it in psychological prominence. Seven *S*s requested and were allowed a change of target number during the course of their test session, and that fact was recorded at the time. When the data from these *S*s are omitted, the significance of the observed page decline is increased.

During the first, or cup-throwing, part of the experiment, *E* stood by the die tray and called out the die faces as he recorded them in pencil. Erasure on data pages was forbidden. The *S* was instructed to allow the dice to lie until *E* called the faces. A single die escaping from the tray or cocking at the tray wall invalidated the throw.

During the second, or machine-throwing, part of the session, *E* and *S* were seated on opposite sides of a table which supported the die-throwing machine. The *E* operated the machine and recorded the data. The *S* was not allowed to touch the machine. The data were recorded manually, *E* calling the die faces as he wrote them down, just as in the cup throwing. In a few cases where *S* objected to *E*'s calling, the machine dice were recorded silently.

Each throw, or half turn, of the cage was serially numbered by a non-resettable Veeder counter, permanently connected to the cage and automatically advanced by its motion. The serial number of the throw was recorded on the data sheet beside the die face entries at the beginning and end of each block of data within the data page.

The organization of the data page was the same for cup throwing and machine throwing, there being one page of cup throwing and two pages of machine throwing. The uppermost faces of the fallen dice were recorded on the data page in blocks of 12 throws (24 dice). There were three data blocks in the upper half page, and three in the lower, making 144 dice per page and 432 dice per *S*. The plan of the experiment required that after each half page the throwing be interrupted to tally the total number of target hits in each of the three data blocks.

In the case of the machine the die faces were also recorded photographically by the automatic, single-frame operation of a modified, 16-mm. motion picture camera. The camera solenoid only was switched off at the end of each upper half page, while counting score. The die cage

[2] To save journal space, 10 pages of supplementary material have been deposited with the American Documentation Institute. Order Document No. 4633 from ADI Auxiliary Publications Project, Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm or $1.25 for photocopies.

itself was stopped only at the bottom of each page.

Overlapping controls were provided against selection of data by improper starting or other mishandling of the equipment. These involved a standardized starting procedure, interlocking of illumination and camera operation, provision of a shutterless timing exposure at the edge of each film frame, and a concealed Esterline-Angus Operation Recorder connected to the camera solenoid.

A detailed description of the apparatus, along with these controls and other precautions such as the maintenance of a photographic film log, the X-ray examination of the dice, and the serial numbering of all data sheets in advance, will be found in an auxiliary publication (Footnote 2).

The analyses presented in this paper have been based upon the hand-written records and have been independently validated by the photographic film. The data-block scores were checked, transferred to IBM punch cards, self-printed, proofread, and independently totaled on two kinds of IBM machines. The original data sheets were then microfilmed. After that, a sample of the photographic records was chosen consisting of approximately 50% of all machine-thrown data involved in the statistically significant page decline reported in the next section. The die faces were transcribed from the film and analyzed by a procedure controlling against motivated errors involving cocked dice, etc. The details have been given auxiliary publication (see Footnote 2).

The results can be summarized by saying that when all photographic film corrections are applied, the observed machine-thrown page decline is increased by a negligible amount over the (manual record) figure quoted in this paper. The accuracy of the hand recording is indicated by the fact that among the more than 18,000 die faces transcribed from the film and checked, no instance was found where a nontarget die had been misrecorded as a target die or vice versa.

Results

Earlier investigators had found two extra-chance statistical effects while throwing dice in a presumably chance situation. The present experiment was conceived primarily as a search for these same effects: namely, a total deviation of target faces above chance expectation, and a decline in the frequency of target faces.

All of the data of this experiment are represented in Table 1, which gives the totals of the scores for each data-block position on the three data pages. The total number of die faces in this experiment was 169,776. The number of target faces was less than theoretical expectation by 91. This deviation is not significant. Thus, the first of the two statistical effects consituting psychokinesis was not found.

The decline effect is not as easily analyzed as the total deviation. It is necessary to specify both the nature of the decline and the statistical method to be used in its evaluation. A decline in scoring rate might extend throughout the work of each *S*, or might exist within the data page, the "set," or the data block. (The "set" is the amount of data taken without interruption. In this experiment the set is the half page of data.) Likewise, there are a number of accepted statistical methods that might be

TABLE 1

Total Number of Successes (Target Faces) for Each Data-Block Position of the Three Data Pages

(The number of *S*s is 393. The number of die-face trials is 9432 for each position. Nominal *P*-value = ¼.)

Half page	Page 1 Cup-Thrown			Page 2 Machine-Thrown			Page 3 Machine-Thrown		
Upper	1603	1532	1570	1580	1617	1550	1661	1587	1545
Lower	1533	1542	1530	1600	1500	1543	1612	1590	1510

applied to a decline to show its probability of occurrence by chance alone. If the question of selection of the analysis is to be avoided, the method of the analysis must be clearly and uniquely sanctioned by prior use in the research in which the decline effect was discovered. Fortunately, in this case it is possible to choose the proper method of analysis without ambiguity.

Although several methods of decline analysis have appeared in the literature, some of these have been tailored to fit the individual experiment. Only two methods have been so widely used as to deserve consideration here. These are the quarter distribution of the page (**3, 8**) and the quarter distribution of the set (**9**). Of these, the quarter distribution of the page has been more widely employed. This method of analysis was originally chosen for its nearly universal applicability and for its sensitivity to the decline effect. It was applied to more than a half-million die throws available at the time of publication of the first reports on the psychokinetic effect (**8,** p. 22). Upon this analysis, more than any other, the case for psychokinetic declines may be said to rest.

The quarter division of the page is accomplished by halving the data page with horizontal and vertical lines. Where the vertical line cuts a data column, as in the case of the present experiment, that column is to be omitted (**1,** pp. 252 and 256; **5,** p. 195; **6,** p. 292; **9,** p. 257; **10,** p. 155). Quoting from the first paper on the quarter-distribution analysis: "It will be in the measure of the decline shown by the diagonal distribution that the significance of a quarter-distribution will be judged" (**8,** p. 25).

Thus, the statistical problem is that of evaluating the difference between the score totals in the upper left and lower right quadrants, that is, between the first and last data blocks on the data page. This computation is done as follows, using the numbers from Table 1: 1603 − 1530 + 1580 − 1543 + 1661 − 1510 = 261 die faces. The binomial *SD* of this composite difference is 88.7, regardless of whether the observed or theoretically expected values of p and q are used. The same critical ratio, 2.94, is also obtained if one employs an arc-sine transformation of the total scores under which the variance is independent of dice bias.

Alternatively, one may apply the null hypothesis to the sum over all pages of the first-last data-block score difference on the page. The observed distribution of first-last data-block score differences approximates normality, and may be used to estimate the variance of the sum of the differences. The *SD* of the difference in score totals based upon these observed first-last data-block score differences is 90.5. The resulting critical ratio is 2.88 which is 2% smaller (yielding a larger probability) than the ratio based upon the binomial hypothesis, and will therefore be used here. Since the number of *df*'s is large, the t variate reduces to the normal variate. The probability by chance alone of a decline as great as that shown is .0020.

Special interest attaches to the last two out of three pages by each *S* because these were made under the more rigorous conditions, with the motor-driven cage and recording camera. Here, the critical ratio of the quarter distribution of the page is 2.60, based upon the variance of the observed data-block score differences.

The preceding analyses, dictated by precedent, establish the extra-chance character of the data by the customary standard of significance. How-

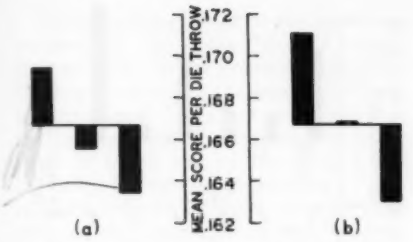

FIG. 1. The target score pattern for the combined half page. Deviation of the mean score per die throw from the theoretical mean chance expectation of $P = \frac{1}{6}$, versus the data-block position on the half page: (*a*) for all half pages of data combined; (*b*) for all machine-thrown half pages combined.

ever, the half page, rather than the page, is the natural psychological unit in the present experiment. In Fig. 1 the half pages have been combined to show their average internal deviation from theoretical expectation. The difference between the first and third data block for all data, as shown in Fig. 1a, has a ratio of 2.72 relative to the theoretical *SD* of that difference. When the machine-thrown half pages are considered separately, as shown in Fig. 1b, this critical ratio is 2.98.

If the variance among the 18 positions of Table 1 is partitioned in the usual way, the critical ratio of 2.72 just cited for Fig. 1a corresponds exactly to the linear component of the variance within the half page. This is the only component of variance that reaches statistical significance.

The chronological accumulation of the half-page decline is plotted in Fig. 2 by groups of 10 *S*s. Here it is shown that the score difference of 341 die faces which has the critical ratio of 2.72 just mentioned, was accumulated more or less continuously throughout the experiment. Had this curve fluctuated more violently in reaching its terminal value, it would indicate that the individual *S* variances were appreciably larger than the theoretical binomial value, *npq*. That such was not the case has been shown by exact calculation in the case of the quarter-distribution analysis of the page.

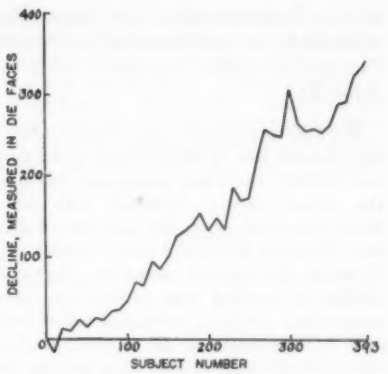

FIG. 2. Chronological accumulation of the combined half-page decline: First and third data-block score difference versus the *S* number, computed for every 10 *S*s. The final decline of 341 die faces corresponds to the difference between the first and third bars of Fig. 1a.

The decline pattern within the half page is given for each of the three *E*s in Fig. 3. The ratio of the difference between the first and third bar to the theoretical *SD* of the difference is, respectively, for each of the three *E*s: .5, 2.5, and 2.0. This variation among *E*s is not statistically signifi-

FIG. 3. Pattern of mean target scores for all half pages, combined by *E*: Experimenter *M*, 173 *S*s; Experimenter *S*, 132 *S*s; Experimenter *P*, 88 *S*s.

cant. Consequently, these data offer no basis for a conclusion as to whether the decline effect occurred with all three *E*s.

Because the significance of the present experiment lies in the decline of scoring rate within the data page, and because the same target number was used throughout the page, the question of dice bias toward a particular face is irrelevant. It might be argued, however, that the decline in scoring was caused by some unspecified physical influence called into play by the start of die throwing. Included here might be such things as temperature changes, electrostatic forces, and condensed vapor layers on the dice. This counterhypothesis would presume that, whatever the physical effect might be, it was modified as the dice approached an equilibrium operating condition, and in this way gave rise to the scoring decline. This idea will now be examined in detail.

The record of how the dice fell exists independently of the particular target number for which *S* wished. That is to say, the way in which the dice were thrown and recorded was in no way different regardless of whether *S* wished for 1's, for 2's, or 3's, etc. One could, for example, perform a large number of imaginary experiments merely by rescoring the same available data according to a new sequence of target numbers. In particular, one might set up a random sequence of targets (one target per *S*) and calculate the resulting "decline effect." It can be shown mathematically that for a random sequence of targets the decline effect must (on the average) equal zero regardless of how biased or nonrandom the dice behavior may have been.

The basic idea here is that since a die can fall in only six mutually exclusive ways, a tendency for one face to appear in any cyclic pattern must be exactly compensated by the tendency of all other faces collectively to appear in the inverse pattern. If, therefore, from *S* to *S*, the target number wanders randomly between one and six, all systematic scoring

FIG. 4. The rate of page decline compared for target and nontarget numbers, for all 386 *S*s who used a single target number within their experimental session.

patterns will cancel out except those that are linked to the target-number sequence. This type of experiment is automatically self-controlling against extraneous physical influences, provided the target sequence is random.

In the present experiment, the target numbers that were actually used form an essentially random sequence—with one exception: some target numbers were used more often than others.[3] This raises a logical possibility that the observed decline effect was the result of some uncomprehended, but ordinary, physical cause associated with the die faces that were aimed at more often, and that the concomitant, compensating incline effect occurred on those die faces that were aimed at less often.

However, when the results obtained are separated according to target number, it is found that a decline effect occurred for *all* targets. Thus, the decline effect cannot be explained by the foregoing counterhypothesis. At most it can be slightly reduced (or increased) in size. By a further analysis of the data it is possible to determine quantitatively how much the decline effect can be changed under the assumption that some spurious causal factor is at work in the manner above described.

[3] When the data are divided according to whether the target number was chosen by *S* or taken from *E*'s list of numbers, the rate of page decline for all pages is .0100 die faces per die pair for the 112 *S*s who chose targets and .0089 for the 281 *S*s who accepted targets from the list. The difference is not significant.

To make this evaluation the data were rescored to show the "decline effect" that occurred on the number *one* face for those trials in which the number *one* was *not* the target. This was compared with the decline effect that occurred for those trials in which *one was* the target, and similarly, for the targets *two* to *six*. The results are shown in Fig. 4, for more than 98% of the data used in the "quarter division of the page" which yielded the previously quoted probability figure of .002. The only deletions are of those seven *S*s who, as mentioned earlier, requested a change of target during the experimental session.

These results can be used to evaluate the starting effect counterhypothesis in the following way. The target and non-target declines for face number *one* are combined to determine a total rate of decline which, lacking further information, one must assume to be typical of face number *one* under the assumption that the psychokinetic effect does not occur. The rate of decline when someone is wishing (i.e., the target decline) is found to be higher than the total rate of decline. Only the difference in these rates of decline is counted as evidence of psychokinesis. When this calculation is made for all of the faces from *one* to *six*, the total target decline is reduced from 263 to 261.5 faces, an entirely negligible decrease. (The variation shown among the several die-face numbers in Fig. 4 is not statistically significant.) Further details are given in an auxiliary publication (see Footnote 2).

Summary

Approximately 170,000 dice were "thrown" by 393 previously untested persons to determine whether dice are influenced by "wishing." Two-thirds of these dice were automatically photographed within a totally enclosing motor-driven cage operated by *E*. The total score was not significantly different from chance expectation. However, in the manner anticipated from the work of earlier *E*s, the rate of scoring declined over the data page with a chance probability of .002.

References

1. McMahan, E. A. PK experiments with two-sided objects. *J. Parapsychol.*, 1945, **9**, 249–263.
2. McMahan, E. A. A PK experiment with discs. *J. Parapsychol.*, 1946, **10**, 169–180.
3. Pratt, J. G. A reinvestigation of the quarter-distribution of the (PK) page. *J. Parapsychol.*, 1944, **8**, 61–63.
4. Pratt, J. G. Rhythms of success in PK test data. *J. Parapsychol.*, 1947, **11**, 90–110.
5. Pratt, J. G. Restricted areas of success in PK tests. *J. Parapsychol.*, 1947, **11**, 191–207.
6. Rhine, J. B. The PK effect: early singles tests. *J. Parapsychol.*, 1944, **8**, 287–303.
7. Rhine, J. B. The problem of psi-missing. *J. Parapsychol.*, 1952, **16**, 90–129.
8. Rhine, J. B., & Humphrey, B. M. The PK effect: special evidence from hit patterns: I. Quarter distributions of the page. *J. Parapsychol.*, 1944, **8**, 18–60.
9. Rhine, J. B., & Humphrey, B. M. The PK effect: special evidence from hit patterns: II. Quarter distributions of the set. *J. Parapsychol.*, 1944, **8**, 254–271.
10. Rhine, J. B., Humphrey, B. M., & Pratt, J. G. The PK effect: special evidence from hit patterns: III. Quarter distributions of the half-set. *J. Parapsychol.*, 1945, **9**, 150–168.

(Received November 12, 1954)

Journal of Experimental Psychology
Vol. 50, No. 4, 1955

EFFECT OF ADDITIONAL SPATIAL REFERENCES ON DISPLAY–CONTROL EFFICIENCY [1]

W. D. GARVEY AND L. L. MITNICK

Naval Research Laboratory [2]

In an attempt to develop a systematic theory of display-control design, several studies (**1, 2, 3**) have demonstrated the importance of properly matching stimulus and response sets to obtain optimum performance from the operator of a display-control system. These studies showed that performance was best when display and control were matched so that an isomorphic relationship existed between stimulus and response sets; i.e., the spatial arrangement of stimulus elements on the display was identical to that of the response elements on the control. The purpose of the studies reported here was to determine if the efficiency of a display-control system can be increased independently of changes in the isomorphic relationship between the S-R sets.

Experiment I

Figure 1 shows the four display-control systems used in this experiment. They are made up of display panels of neon lights, push-button controls, and the center lines; these systems are the same as those used in an earlier experiment (**2**) except for the additional center lines shown on this figure. Since, in the earlier experiment, systems without lines were indicated by letters C, D, E, and F, the present systems *with lines* become, respectively, C′, D′, E′, and F′.

In the earlier experiment (**2**) analysis of response times to the different stimuli on System C revealed a sharp increase in response times from the periphery to the center of the matrix, i.e., response times to stimuli arising in the center of the display were approximately three times as great as those in the periphery. A similar analysis of response times to stimuli on System D revealed a more complex pattern, since two lights constituted a stimulus element. In general, it was found that the farther a stimulus element was located from either the top or bottom of a column, the longer the resulting response time. However, the maximum differences between response times to stimulus elements located in the center and those located at the ends of the columns were not as large as the differences found between the peripheral and central response times for System C. Analyses of response times for Systems E and F did not reveal either of these patternings. In general, response times to signals on Systems E and F were of the same order of magnitude as those found for the most centrally located lights for System C.

These response-time patterns indicated that in Systems C and D an isomorphic relationship between stimulus set and response set permitted *S*s to use certain spatial references in responding to some of the stimuli:

[1] The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the views of the Navy Department or the naval services at large.

[2] The writers wish to acknowledge the valuable suggestions made by Drs. W. B. Knowles and F. V. Taylor in setting up the experiment and interpreting the data. Also they wish to express their appreciation to Miss Jean B. Henson and Mr. E. P. Newlin, who assisted in collecting and analyzing the data.

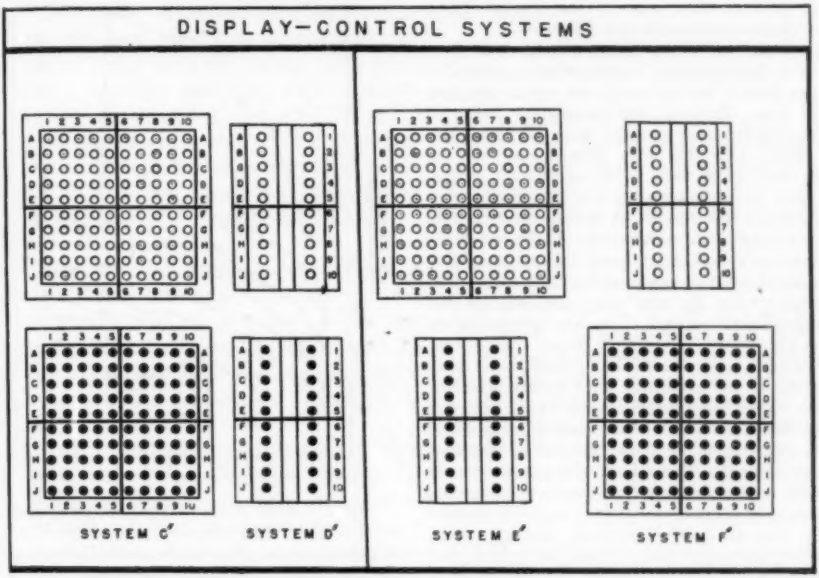

FIG. 1. Displays and controls. The hollow circles represent the display lights and the filled circles represent the control buttons. The horizontal and vertical lines represent the additional lines placed on the units for the experimental group.

e.g., the sides of the matrices in System C and the ends of the columns in System D. Longer responses were made to stimulus objects located remote from these spatial references. It would appear that internal interference was greatest among those S-R elements which were not situated near the spatial anchors. Thus, one way of reducing this interference would be to provide additional references at the location of the greatest interference. Additional references were provided in the present experiment by adding horizontal and vertical center lines so that Systems C′ and D′ are formed as shown in Fig. 1. It was hypothesized that the net result would be an overall increase in the efficiency of operating Systems C′ and D′.

An isomorphic relationship between display and control and the resulting spatial references do not exist in Systems E and F; *Ss* have to transform information from one coordinate system to another. Thus it was hypothesized that there would be no appreciable reduction in the amount of internal interference nor any overall improvement in performance by the addition of center horizontal and vertical lines.

Method

Apparatus.—Detailed descriptions of the general apparatus have been reported (**2, 3**). Two displays were used: (*a*) a 10 × 10 matrix of 100 lights, and (*b*) two columns of 10 lights. The physical layouts of the controls and the displays were equivalent: (*a*) a 10 × 10 matrix of push buttons, and (*b*) two columns of 10 push buttons. For the experimental group lines were present on the displays and controls; for the control group these lines were not present. The *Ss*' task was to press the control buttons identical

in letter-number code to the display lights. Stimulus presentation was self paced.

Procedure.—Two weeks' practice was given on a display-control system which consisted of two rows of ten lights and two rows of ten push buttons. Previous experimentation indicated that performance on this system was highly correlated ($r = .93$) with performance on the systems to be used in the experiment proper. Thus, it was planned to use the practice performance scores to adjust the experimental data in an analysis of covariance. At the end of the practice period, eight naval enlisted men were divided into two groups of four men each. The experimental *Ss* were given one trial of 100 stimuli a day on each of the four systems shown in Fig. 1 for a period of 26 days. Sixteen randomized sequences of 100 stimuli were used. Systems and sequences were counterbalanced, the orders being systematically varied from day to day. Efficiency was measured in terms of rate of responding (mean number of responses per second). Throughout the experiment *Ss* were reminded before each trial how the system was to be operated and that they should respond as fast as possible without making errors. Error scores were also recorded as a check that *Ss* were not achieving speed at the expense of accuracy. The control group operated under these same experimental conditions but with the reference lines omitted.

From Trials 27 through 36 the lines were removed from the systems which the experimental group had been operating, and this group was required thereafter to perform without benefit of lines. The control group continued to operate the systems during this period as they had done for the first 27 trials.

Results [3]

The effect of adding lines.—The general results are shown in Fig. 2. Using the means of the last five pre-experimental task trials as the adjusting data, analyses of covariance were performed. The differences between each experimental system and its control were tested for significance. From Trials 1 through 26, using only the data for every third trial, the experimental group responded faster on Systems C′ and F′ than the control group did on Systems C and F ($P < .01$). The differences in rate of responding were not significant between Systems D and D′ ($P > .10$) nor between Systems E and E′ ($P > .05$). With all systems, errors were found to average about 3% of the responses and were distributed randomly. It should also be pointed out that the efficiency of systems with lines ranked in the same order as those without lines.

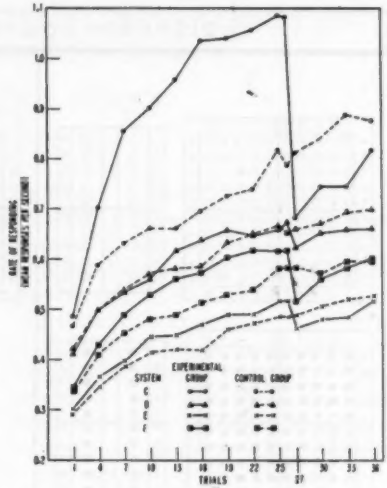

FIG. 2. Rate of responding (mean responses per second) for the display-control systems on the 36 practice trials. One trial consisted of each of the 100 possible stimulus alternatives.

The effect of removing lines.—The data of Trial 27 when subjected to similar analyses of covariance indicated that when the lines were removed from System C′, the experimental group responded at a rate significantly slower than the control group ($P < .01$). Removal of the lines from the other three systems resulted in no significant differences between groups.

[3] A 24-page table giving critical summaries of data as well as the results of the statistical analyses has been deposited with the American Documentation Institute. Order Document No. 4634, remitting $2.00 for microfilm or $3.75 for photocopies.

Analyses of covariance performed on the data for Trials 27 through 36, again using data for every third trial only, showed that the experimental group continued to operate System C at a significantly slower rate than the control group throughout these trials ($P < .05$); no significant differences were established between groups for the other systems.

Removal of the lines resulted in a significant decrease in rate of responding for all systems when the differences between Trials 26 and 27 for the experimental group were tested by the matched t test ($P < .05$).

The analysis of component S-R elements.—The polygraph records for Trials 23 through 26 were analyzed in detail and a mean response time for each S-R element was obtained from this sample of 16 readings. Three-dimensional models such as those presented in the previous study (2) were constructed to show mean latencies of the S-R elements as the heights of bars, where the cells represent specific stimuli identified by the letter-number code. To test for the presence of a pattern each response time matrix was tested for a significant row by column interaction by analysis of variance. These analyses indicated that the patterns for all systems except C′ differed reliably from a flat surface ($P < .05$). To test for differences between the patterns of a system with lines and its control, a matrix of cell differences was constructed and an analysis of variance was performed to determine the significance of the row by column interaction. The pattern differences between Systems C and C′ were found to be highly significant ($P < .01$). Systems D, E, and F when tested against D′, E′, and F′, respectively, revealed no significant differences in the response-time patterns.

Discussion

The results point out two important factors determining the efficiency of a display-control system: (*a*) the isomorphic relationship between stimulus and response sets, and (*b*) the degree to which internal interference among stimuli may be eliminated.

The addition of the center lines had the most beneficial effect on System C. The response-time model for System C indicated that the internal interference is greatest among stimuli arising from the center of the display. By placing lines on the matrices so as to intersect at the middle of the central lights and buttons, the interference among the stimuli was reduced. Adding lines reduced the response latencies to the central signals to a level equivalent to that of the peripheral signals. Thus, without disturbing the isomorphic relationship between stimulus and response sets, the efficiency of System C was improved by providing additional spatial structuring to the panels.

On the other hand, no significant differences were found between the performance of *S*s on Systems D and D′. An examination of the response-time models for Systems D and C revealed that the ratio of response times in the area of maximum interference to the area of minimum interference was 1.4, where as for System C the equivalent ratio was 3.2. This suggests that if the internal interference among stimuli is not large as in System D, then the addition of lines should not furnish spatial anchors of any psychological advantage in reducing the internal interference.

The relationship between display and control of System E is such that there are no clear-cut spatial relationships between stimulus and response sets. Thus, adding lines would not be expected to furnish additional spatial references on this system. The same reasoning should hold true for System F. However, System F *was* improved by the addition of lines. Since this system permits no simple spatial relationship between stimulus and response sets, this result suggests the appropriateness of an hy-

pothesis which differs from the original hypothesis that reference lines will tend to increase the efficiency of isomorphic display-control systems and will not increase efficiency in the nonisomorphic systems. When it is noted that the one thing in common between Systems C′ and F′ is that they use a matrix for controls, it is suggested that the factor which interacts with reference lines is not isomorphism between displays and controls, nor is it reference lines in the stimulus component (as in E′); rather the results suggest that the introduction of lines may serve to reduce the interference of control elements when such elements are arranged in a matrix.

Effect of removing the lines.—The removal of lines on Trial 27 caused a significant disturbance of performance on all systems ($P < .05$), but it is evident that disturbance was greatest when System C′ was changed to System C. This would indicate that *S*s were relying on the lines to operate all systems to some extent. With Systems C′ and F′ such reliance resulted in better performance than was obtained with the control group, while with Systems D′ and E′ no significant difference in performance between the two groups was found. However, removal of the lines disturbed performance to the extent that it dropped significantly below the mean level of performance of the control group only on System C′. It is significant to note that, when the lines were removed from System F′, experimental *S*s began to respond slower as compared to the previous trials, but not significantly slower than the control *S*s.

It appears that *S*s may have developed a specific method for coding certain S-R elements on System C′ which was dependent upon the presence of lines. When the basis upon which this method of coding depended was removed, *S*s had to learn a new code for these S-R elements. The fact that the experimental group remained poorer than the control group for the remaining 10 trials would indicate a difficulty in abandoning the old and adopting a new method of coding. Furthermore, that this was not the case for System F would lend support to the inference that the lines did not alter the operator's method of coding with this system; rather they facilitated in some manner a method similar to that which control *S*s used without benefit of lines.

Experiment II

In Exp. I the addition of two lines on the display and control of System C materially improves performance. Observation of the response-time model for System C′ revealed that it was essentially flat. It would appear that adding two lines has reduced the internal interference among the centrally located stimulus elements so that *S*s respond to all stimuli with equivalent difficulty. Thus, it would be predicted that two reference lines are sufficient for reducing the internal interference among stimulus elements on this system, and no further benefit would be gained by adding more than two lines. Exp. II was conducted to test this prediction.

Method

In general the apparatus and procedure was identical to that of Exp. I. The matrix display-control system and four modifications shown in Fig. 3 were used in this experiment. An 8×8 matrix, instead of the 10×10 of Exp. I was used to permit symmetrical placement of the reference lines.

Five naval enlisted men, not used in Exp. I, were given two weeks of practice on the same pre-experimental task given the groups in Exp. I. Each *S* was then presented 64 stimuli (each of the possible S-R alternatives) a day on each of the five systems shown in Fig. 3 for a period of 23 days. Systems and sequences were counterbalanced and varied systematically from day to day.

Results[4]

The effect of additional lines.—The results are presented in Fig. 4. Matched *t* tests performed on the data for the last eight trials indicated

[4] See footnote 3.

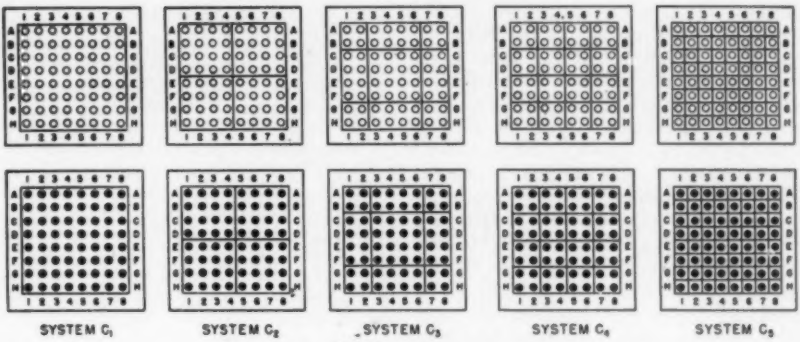

FIG. 3. Display and controls in Exp. II.

that: (*a*) performance on Systems C_1 and C_5 was equivalent ($P > .05$), and (*b*) performance on Systems C_2, C_3, and C_4 was also equivalent ($P > .05$), but (*c*) performance on Systems C_1 and C_5 was significantly poorer than that on Systems C_2, C_3, and C_4 ($P < .01$).

The analysis of component S-R elements.—Analyses of the response-time patterns similar to those of Exp. I indicated that the patterns for Systems C_1 and C_5 were reliably different from a flat surface ($P < .05$), but that the difference between the two surfaces was not significant ($P > .05$). Patterns for Systems C_2, C_3, and C_4 were not found to differ significantly from a flat surface ($P > .05$) and the differences between these patterns were attributable to chance ($P > .05$).

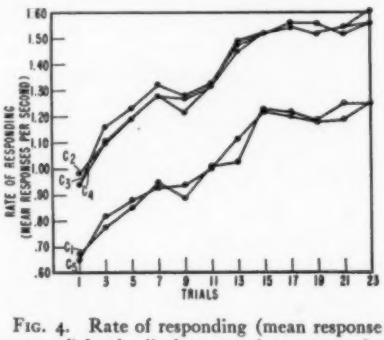

FIG. 4. Rate of responding (mean response per second) for the display-control systems on the 23 practice trials. One trial consisted of each of the 64 possible stimulus alternatives.

DISCUSSION

From Fig. 4 it can be seen that there was no further improvement in performance caused by adding more than two lines. It may be inferred that for a system of this size two lines are sufficient to reduce internal interference among the stimuli in the center to a minimum. When all the lights and buttons are surrounded by lines, performance became as poor as when no lines were present at all. These additional lines prevented *S*s from using spatial references other than the ones used in operating System C_1. That there were no differences either in rate of responding or the response-time patterns would lend support to this reasoning.

These studies were conducted in an effort to obtain a better understanding of the factors underlying efficient display-control relationships. Display-control relationships which proved to be most efficient have all exhibited an isomorphic relationship between stimulus and response sets. These isomorphic relationships between sets engender spatial cues, which facilitate *S*'s performance in identifying S-R elements. The ef-

ficiency of a system can be further enhanced by providing additional spatial references to those S-R elements where the original references are least effective and the interference among stimuli is greatest. It is likely that the human operator uses certain configurational processes which are effective in organizing spatial groupings of lights and buttons and in providing references which are not directly present in the structural aspects of the system. Thus, although *S*s may rely more on the structural references with System C_4 than they do with System C_2, the two lines in System C_2 seem to supply sufficient cues to the operator to make configurations on C_2 equivalent to C_4 with the result of no difference in performance on the two systems. The results, therefore, support the prediction made from Exp. I, i.e., since the addition of two lines reduces the internal interference among stimuli to a level where all S-R elements are equivalent in difficulty, no further benefit is gained by increasing the number of referents. Furthermore, the data indicate that the addition of an excess of spatial referents significantly interferes with performance. In terms of display-control design it appears that the benefit to be gained by adding spatial referents such as lines, markers, indices, etc. is determined by the degree of interference among stimuli, and that it is equally poor design to provide referents in excess of those necessary to reduce the interference to a minimum as it is to neglect the provision of referents adequate to reduce the interference.

Summary

Two experiments were conducted to determine the effect of additional spatial references on display-control efficiency. The purpose of Exp. I was to determine if the efficiency of a display-control system could be increased independently of changes in the isomorphic relationships between S-R sets by the addition of spatial references on each display and control in the area of the greatest interference. A measure was obtained of amount of internal interference among the stimuli associated with each display-control system. In Exp. II, the number of additional spatial reference lines was varied. The results indicated:

1. The degree of isomorphic relationship between S-R sets is an important determinant of display-control efficiency.
2. The efficiency of a display-control system can be further enhanced by providing the operator with additional spatial references on the system where the internal interference among stimuli is greatest.
3. The addition of spatial references, after internal interference had been reduced to a minimum, does not further enhance performance.
4. The addition of an excessive number of spatial references significantly deteriorates performance.

References

1. Fitts, P. M., & Seeger, C. M. S-R compatibility: spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.
2. Garvey, W. D., & Knowles, W. B. Response time patterns associated with various display-control relationships. *J. exp. Psychol.*, 1954, **47**, 315–322.
3. Knowles, W. B., Garvey, W. D., & Newlin, E. P. The effect of "speed" and "load" on display-control relationships. *J. exp. Psychol.*, 1953, **46**, 65–75.

(Received November 15, 1954)